



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 1997

Forschungsmethoden

Widmer, Thomas ; Binder, Hans-Martin

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-158567>

Book Section

Published Version

Originally published at:

Widmer, Thomas; Binder, Hans-Martin (1997). Forschungsmethoden. In: Bussmann, Werner; Klöti, Ulrich; Knoepfel, Peter. Einführung in die Politikevaluation. Basel: Helbling Lichtenhahn, 214-255.

KAPITEL 12: FORSCHUNGSMETHODEN

Thomas Widmer, Hans-Martin Binder¹

Jede Evaluation stützt sich auf Informationen, die bei den Forschungsteams zunächst nicht vorhanden sind. Das vorliegende Kapitel befasst sich mit der Frage, wie diese Materialien und Daten, die für die Beurteilung der Wirkung einer politischen Massnahme unentbehrlich sind, erhoben und ausgewertet werden sollen. Es geht somit um die Forschungstechniken.

Die Auswahl der Forschungsv erfahren steht meist in einem Zusammenhang mit der erkenntnistheoretischen Position der Forscherinnen und Forscher. Im Abschnitt 12.1. unterscheiden wir diesbezüglich das empirisch-analytische Paradigma von interpretativ-hermeneutischen Ansätzen. Die Darstellung der eigentlichen Forschungstechniken ist auf zwei Abschnitte verteilt. Zunächst werden (12.2.) die qualitativen und anschliessend (12.3.) die quantitativen Verfahrensweisen erörtert. In beiden Abschnitten geht es um die Erfassung der Materialien und Daten, um ihre Auswertung sowie um die Fragen der Gültigkeit und der Verlässlichkeit der daraus gezogenen Schlüsse. Schliesslich enthält das Kapitel einige Hinweise auf die Probleme der Darstellung der Ergebnisse von Evaluationen.

12.1. FORSCHUNGSPARADIGMEN

In den Sozialwissenschaften gibt es seit langem eine Kontroverse zwischen zwei grundlegenden erkenntnistheoretischen Positionen. Die eine ist an den Naturwissenschaften, die andere an den Geisteswissenschaften orientiert. Die Auseinandersetzung zwischen diesen beiden Richtungen hat sich schon früh auch auf die Evaluationsforschung übertragen (Campbell 1979). Es drängt sich deshalb auf, an dieser Stelle auf den damit verbundenen Methodenstreit kurz einzugehen.

Das an den Naturwissenschaften orientierte *empirisch-analytische Paradigma* bildet in der Evaluationsforschung immer noch den Mainstream. Es geht von einer rationalistischen Sicht der sozialen Welt aus. Diese kann in Teilerscheinungen oder Elemente aufgeschlüsselt werden, zwi-

schen denen Zusammenhänge bestehen. Diese werden durch die Bildung von Variablen erfasst. Aufgrund von theoretischen Annahmen werden Hypothesen postuliert, die es zu überprüfen gilt. Können die Hypothesen nicht falsifiziert werden, so gelten sie vorläufig als gesichert. Gültige Hypothesen werden zu neuen Theorien zusammengefügt, die es erlauben, die soziale Welt zu erklären. Die Evaluationsforschung, die sich diesem Ansatz verpflichtet fühlt, neigt dazu, mit experimentellen oder quasi-experimentellen Untersuchungsdesigns zu arbeiten. Sie stützt sich dabei auf deduktiv hergeleitete Wirkungsmodelle, die sie in der Wirklichkeit einem rigorosen Test unterzieht.

Im Gegensatz dazu steht das geisteswissenschaftliche *interpretativ-hermeneutische Paradigma*. Dieses geht davon aus, dass die soziale Welt nur in ihrer Totalität erfassbar ist. Zu deren Entschlüsselung ist ein verstehendes Vorgehen erforderlich. Im strengen Sinne gibt es keine objektive gesellschaftliche Wirklichkeit, sondern lediglich verschiedene subjektive Rekonstruktionen davon. Erkenntnisse sind deshalb letztlich nur über Kommunikation zu gewinnen. Die Interpretation dieser perspektivenabhängigen Sichtweisen erfordert zwar Vorkenntnisse und Vorstrukturierungen des Untersuchungsfeldes; sie zwingt indessen auch zu einer Offenheit gegenüber neuen Informationen, welche die ursprünglichen Annahmen im Laufe des Forschungsprozesses verändern können. In der Evaluationsforschung bedeutet dies, dass vorwiegend das Design der Fallstudie gewählt wird. Dieses erlaubt es, möglichst alle mit einer Massnahme zusammenhängenden Erscheinungen in dichter Weise zu beschreiben und so die Grundlagen für die Beurteilung von Wirkungen bereitzustellen.

Die beiden geschilderten Ansätze haben selbstverständlich auch gewisse *Affinitäten* zu den in den kommenden Abschnitten vorzustellenden *Forschungstechniken*. So sind Untersuchungsdesigns, die auf dem empirisch-analytischen Ansatz basieren, mindestens teilweise auf quantitative Daten angewiesen. Umgekehrt sind mit dem interpretativ-hermeneutischen Paradigma fast ausschliesslich qualitative Forschungsv erfahren verbunden.

¹ Hans-Martin Binder: Ausschliessliche Verantwortung für den Abschnitt 12.2.; für die übrigen Teile liegt die alleinige Verantwortung bei Thomas Widmer.

Abbildung 7 Verbindung von Paradigma und Forschungsverfahren

Paradigma	Forschungsverfahren	
	quantitativ	qualitativ
empirisch-analytisch	Normalfall	gelegentlich
interpretativ-hermeneutisch	selten	Normalfall

Unsere Übersicht will indessen zeigen, dass die Verbindungen von "empirisch-analytisch" mit "quantitativ" sowie von "hermeneutisch" mit "qualitativ" lediglich den Normalfall darstellen, dass sie aber keineswegs zwingend sind. Es wird somit keineswegs postuliert, dass quantitative Verfahren der Datenerhebung und -auswertung nur im Rahmen des empirisch-analytischen Ansatzes zur Anwendung gelangen können, auch wenn in der Forschungspraxis Exponenten des interpretativ-hermeneutischen Paradigmas nur selten mit quantitativen Daten arbeiten. In der Praxis ist es umgekehrt recht häufig, dass empirisch-analytisch orientierte Forscherinnen und Forscher auch qualitative Daten in ihre Untersuchungen einbeziehen. Keine Kombination ist auszuschließen. Wenn demnach im folgenden für die Darstellung der quantitativen und der qualitativen Verfahren jeweils die Kombination mit dem Paradigma gewählt wird, das den Normalfall bildet, so geschieht dies nicht aus systematischen, sondern aus didaktischen Gründen und zur Vereinfachung der Darstellung.

12.2. QUALITATIVE VERFAHREN

Die in diesem Abschnitt diskutierten Forschungstechniken unterstehen einem hermeneutischen, interpretativen Paradigma. Das entsprechende Denkmodell geht davon aus, dass Menschen zwar in einem von aussen beschreibbaren kulturellen, sozialen, politischen und wirtschaftlichen Umfeld handeln, dass sie aber stets ihre Situation und ihre Interaktionen deuten müssen. In Anlehnung an die Theorien des Symbolischen Interaktionismus (siehe dazu Blumer 1976) und der Ethnomethodologie fordert Wilson (1976), dass jede soziale Interaktion als interpretativer Prozess

aufzufassen sei. Menschen müssten jede soziale Situation und jeden Prozess in ihrem eigenen Umfeld für sich deuten, müssten ihre eigenen Rollenzuschreibungen erfassen, und ihre Verhaltens- und Handlungsoptionen beurteilen (vgl. Mayring 1993). Methodologisch kann diesem interpretativen Paradigma mit einem hermeneutischen Ansatz entsprochen werden. Dabei werden Texte im Sinne des Verstehens interpretiert: "Hermeneutisches Verstehen bezieht sich auf das Erfassen menschlicher Verhaltensäusserungen und Produkte" (Lamnek 1988: 65).

Hermeneutische Interpretation ist indessen keinesfalls eine regellose "Freistil"-Methodologie. Dieser sozialwissenschaftliche Verstehensprozess hat sich auch an gängigen Güte-Kriterien wie Validität und Reliabilität zu messen. Die Philosophen Schleiermacher (1768-1834) oder Dilthey (1833-1911) haben unter dem Begriff des "hermeneutischen Zirkels"² entsprechende wissenschaftliche Standards gesetzt (Lamnek 1988: 67 f.). Hermeneutisch interpretative Forschungsansätze haben demzufolge im wesentlichen zwei Prinzipien zu genügen: dem Prinzip der Offenheit und dem Prinzip der Kommunikation.

Das Prinzip der Offenheit beinhaltet als Forderung, "den Wahrnehmungstrichter empirischer Sozialforschung so weit als möglich offen zu halten, um dadurch auch unerwartete, aber dafür umso instruktivere Informationen zu erhalten" (Lamnek 1988: 22). Diese Forderung hat zur Konsequenz, dass theoretische Strukturierungen und Hypothesen, wie auch methodische Verfahren im Verlauf des Forschungsprozesses erweitert, modifiziert und revidiert werden können. Diese Offenheit sowohl gegenüber den Untersuchungspersonen und ihren Kontexten als auch gegenüber den angewendeten Methoden zeigt, dass diese Art qualitativer Sozialforschung eine starke "Explorationsfunktion" aufweist: "Qualitative Sozialforschung legt das Schwergewicht auf die explorierende Felderkundung, die bei Anwendung standardisierter Techniken ohne vorherige Erkundungsphase meist sträflich vernachlässigt wird." (Lamnek 1988: 22; nach Hoffmann-Riem 1980)

² Es werden zwei hermeneutische Zirkel unterschieden (Lamnek 1988: 260):

- I. Ein zunächst rudimentäres Vorverständnis ist zum Verstehen eines Textes und dessen Auslegung notwendig. Das durch das Textverstehen angeeignete Wissen trägt zur Erweiterung und Korrektur des Vorverständnisses bei, womit sich der Text wiederum besser verstehen lässt.
- II. Die Einzelelemente eines Textes sind nur aus dem Gesamtzusammenhang verständlich, der sich wiederum nur aus den Einzelteilen ergibt."

Gemäss dem Prinzip der *Kommunikation* ist davon auszugehen, dass Evaluatorinnen und Evaluatoren in ihrem Forschungsprozess nicht unabhängig sind von den Daten und Informationen, die sie durch die an der Evaluation beteiligten Akteure erheben. Dieses wechselseitige Aufeinanderbezogenheit von Forscherinnen und Forschern und ihrem Evaluandum ist konstitutives Element im Rahmen eines interaktiven Forschungsprozesses. Daten sind nur über Kommunikations- und Verstehensprozesse zu gewinnen. Damit unterliegen sie aber auch subjektiven Deutungen, die im Ablauf bestimmter Interaktionsprozesse erfolgen: Akteure strukturieren und deuten mit ihren Definitions- und Interpretationsleistungen die von ihnen erfasste Wirklichkeit. Demzufolge ist die dargestellte Sicht der Wirklichkeit stets perspektivenabhängig; und mit dem Wechsel der subjektbezogenen Perspektive ändert sich auch die dargestellte Wirklichkeit (Schütze 1978).

Das Prinzip der Offenheit und das Prinzip der Kommunikation haben vorgehensprägende Implikationen auf die Methodik und den konkreten Forschungsablauf einer Evaluation. Vor allem Guba und Lincoln (1989) haben unter der Bezeichnung "Fourth Generation Evaluation" den Ansatz der interaktiven Evaluation stark geprägt. Dieser weist im übrigen zahlreiche Bezugspunkte zum Konzept lernorientierter Evaluationen auf (vgl. Kapitel 9 und 13.3.).

12.2.1. Interaktive Evaluation

Folgende Konzept-Elemente stehen bei diesem Forschungsansatz im Vordergrund:

Evaluation muss alle beteiligten Akteure ("stakeholders") mit ihren Wert- und Zielvorstellungen in einen Prozess der Aushandlung einbeziehen, um deren unterschiedliche Wert- und Zielvorstellungen und die unterschiedlichen Sichtweisen und Realitätskonstrukte miteinander zu konfrontieren. Ziel ist es, ein möglichst facettenreiches Bild des Evaluations-Gegenstandes zu erhalten.

Diesem Konzept zufolge folgt die interaktive Evaluation in ihrem methodologischen Ansatz einer *konstruktivistischen Sicht der Realität*: Es gibt nicht eine "objektive" Realität. Die Beschreibung der Realität ist vielmehr ein Konstrukt des Beobachters bzw. der Beobachterin und ist das

Resultat von Interaktionen zwischen den Beobachtenden und den Beobachteten (Bierter und Binder 1993: 197).

Die im Rahmen eines solchen Evaluationsansatzes zur Anwendung gelangenden methodischen Instrumente müssen sicherstellen,

- dass alle im Zusammenhang mit dem Evaluations-Gegenstand involvierten Akteure identifiziert werden,
- dass deren Wert- und Zielvorstellungen erfasst werden, und
- dass erfasst wird, wie die Akteure ihre Wert- und Zielvorstellungen mit dem Gegenstand und Kontext der Evaluation in Verbindung bringen.

In einem hermeneutischen Auswertungs- und Interpretationsprozess müssen die verschiedenen Wahrnehmungen und Realitätskonstrukte der beteiligten Akteure verstanden und miteinander konfrontiert werden. Über einen Aushandlungsprozess unter Beteiligung der einbezogenen Akteure muss schliesslich versucht werden, einen Konsens im Verstehen und über die gegebenen Interpretationen zu erreichen, aus dem heraus konkrete Handlungsperspektiven zu entwickeln sind (Guba und Lincoln 1989: 72 ff.). Das Ziel dieses Interaktionsprozesses soll darin bestehen, "die Realitätssicht möglichst vieler Beteiligter adäquat zu repräsentieren, divergente Positionen deutlich zu machen und die Basis für Schlussfolgerungen und Empfehlungen klar zu benennen" (Kraus 1991: 413).

Dieser Evaluationsansatz zielt klar auf den notwendigen Nutzenaspekt jeder Evaluation. Eine noch explizitere Nutzenorientierung umfasst der Ansatz der "Utilization-Focused-Evaluation" von Patton (1978), der die Zielintention einer Evaluation noch weit stärker auf die Erarbeitung und Entwicklung von konkreten Handlungsstrategien und -optionen für einen definierbaren Kreis von Akteuren fokussiert.

Allen Ansätzen der qualitativen, interaktiven Evaluationsforschung ist gemeinsam, aus einem Interaktionsprozess von Wissenschaft und politischer Praxis, auf der Basis erkannter und kommunizierter Realitätskonstrukte neue Handlungsperspektiven zu eröffnen. "So verstanden, wäre Evaluation der Versuch, die Kluft zwischen Wissenschaft und Praxis aufzuheben. Die Diskursivierung der Handlungsprogramme aller Beteiligten, wie sie in der qualitativen Evaluationsforschung stattfindet, könnte dann, konzeptuell integriert in eine organisatorische Praxis, idealerweise kollektive Lernprozesse ermöglichen, mindestens jedoch verfestigte Strukturen aufbrechen und eingefrorene Kompromisse verflüssigen." (Kraus 1991:

415). Solche interaktive Evaluation ist v. a. ein wesentliches Merkmal von Mediationsprozessen (vgl. Kapitel 9).

12.2.2. Gütekriterien qualitativer Evaluationsforschung

Auch im Rahmen des interpretativen Paradigmas haben sich die methodischen Ansätze mit den Anforderungen an die Gültigkeit und Verlässlichkeit wissenschaftlicher Messprozesse auseinanderzusetzen. In der methodologischen Diskussion indessen werden Sinn, Adäquanz und Machbarkeit von Gütetests je nach Methodenansatz sehr unterschiedlich beurteilt.

12.2.2.1. Reliabilität

Eine hohe Zuverlässigkeit der Messinstrumente und der Datenerhebungsverfahren setzt - streng genommen - eine identische Wiederholbarkeit des Forschungsergebnisses zu unterschiedlichen Zeitpunkten und die Unabhängigkeit der Messung von der Datenerhebungsperson voraus. Wenn wir aber davon auszugehen haben, dass die Datenerhebung im Rahmen der qualitativen Evaluationsforschung als soziale Interaktion zwischen dem forschenden und dem zu befragenden Akteur zu verstehen ist, dann ist dieser Prozess einer bestimmten Dynamik unterworfen, und demzufolge durch vielfältigen Merkmalswandel gekennzeichnet. Die Stabilität des Messergebnisses bei wiederholter Durchführung ist deshalb - unter positivistischen Prämissen - nicht gewährleistet. Rossi und Freemann (1993: 230) bezeichnen diesen Umstand als gravierenden Mangel, weil unscharfe Messinstrumente reale Wirkungen verwischen oder verschleiern würden; es wäre demzufolge weniger wahrscheinlich, dass eine erfolgreiche Intervention entdeckt werden könne. Demgegenüber attestiert Lamnek dem interaktiven Evaluationsansatz eine nicht zu unterschätzende Stärke durch den Einbezug subjektiver Anteile in diesem kommunikativen Forschungsprozess: "An die Stelle von Replizierbarkeit von Untersuchungsbedingungen und Forschungsergebnis tritt im interpretativen Paradigma die Betonung der "situativen Kontextgebundenheit" von Datenerhebungs- und Auswertungsergebnissen." (Lamnek 1988: 162).

Auch Guba und Lincoln bezeichnen in ihrem Ansatz der Fourth-Generation-Evaluation die Reliabilität im positivistischen Sinn als nicht adäquat für einen konstruktivistischen Ansatz. Sie schlagen mit dem Begriff der "Dependability" ein paralleles Kriterium zur Beurteilung der Stabilität von

Datenerhebungen und Messverfahren vor. Dieses Konzept der Verlässlichkeit muss gewährleisten, dass Aussenstehende den Forschungsprozess und die im Verlaufe dieses Prozesses getroffenen Entscheidungen und Interpretationen des Evaluators nachvollziehen können: "The technique for documenting the logic of process and method decisions is the dependability audit." (Guba und Lincoln 1989: 242).

12.2.2.2. Validität

Die im Rahmen einer interaktiven und interpretativen Evaluationsstrategie hauptsächlich zur Anwendung gelangenden Methoden (Interviews, Gruppendiskussionen) umfassen zum grossen Teil Interaktions- und Kommunikationsprozesse. Unter wahrnehmungs- und erkenntnistheoretischen Aspekten (Benseler, Hejl und Köck 1980; Varela und Thompson 1992; Watzlawick und Krieg 1992) geht es nicht darum, "objektive" Informationen von Befragten zu erhalten; vielmehr interessiert die Forscher und Forscherinnen das vom befragten Akteur kognitiv strukturierte Feld des Handelns und seiner entsprechenden Umwelt (Kontextbezogenheit) sowie seine Begründungen (seine "impliziten Theorien"). Unter dieser Prämisse ist deshalb jede Äusserung der Befragten eo ipso "gültig" (Bierter, Binder und Rüegg 1985: 49 ff.). Das im positivistischen Validitätsverständnis wichtige Korrespondenz-Kriterium (Übereinstimmung einer Behauptung mit der "objektiven" Welt) wird von einer hermeneutisch und konstruktivistisch orientierten qualitativen Evaluationsforschung abgelehnt. Mit einem solchen Perspektivwechsel von der notwendigen Übereinstimmung mit einer "objektiven" Welt hin zu einer Vorstellung von der sozialen Wirklichkeit findet auch eine Verschiebung von der alles prägenden Beobachtung der sozialen Wirklichkeit hin zu einer Auseinandersetzung über sie statt. Ein solcher Wandel erfordert ein kommunikatives und pragmatisches Validierungskonzept (Kvale 1991: 427 ff.).

In Anlehnung an Volmerg (1983: 124 ff.) gilt es im Rahmen des interpretativen Paradigmas neue qualitative Gültigkeitskriterien zu entwickeln. Drei solche Validierungen stehen im Vordergrund:

Ökologische Validierung basiert auf der Überzeugung, dass gültige Informationen über Forschungsgegenstände und Untersuchungspersonen nur in deren "natürlichem Lebensraum" gewonnen werden können. Dieser natürliche Kontext darf nicht durch künstliche oder standardisierende Elemente eingeengt oder entfremdet werden. Ökologische Validierung beschränkt sich nicht auf den Prozess der Datenerhebung, sondern fordert

auch den Einbezug der Kontextbedingungen in den hermeneutischen Interpretationsprozess.

Bei der *kommunikativen Validierung* werden Interpretationsergebnisse durch erneutes Befragen der Akteure auf Stimmigkeit und Gültigkeit hin überprüft. Die Forscherinnen und Forscher vergewissern sich durch diesen Validierungsprozess, ob sie die befragten Akteure richtig verstanden haben und ob diese ihren Interpretationen zustimmen. Kommunikative Validierung bedeutet demnach "die dialogförmige Überprüfung der Validität von Erkenntnisansprüchen" (Kvale 1991: 429). Sie muss die "Interpretationsgemeinschaft" auf die untersuchten Personen ausdehnen, denn nur so kann die notwendige Zustimmung der Befragten zu den Interpretationen des Evaluators - und damit "die Einbeziehung der Öffentlichkeit in die Interpretationsgemeinschaft" - sichergestellt werden (Kvale 1991: 430).

Argumentative Validierung bedeutet, dass das Forschungsteam seine (theoriegeleiteten) Vorannahmen gegenüber den einbezogenen Akteuren offenlegen muss; seine Interpretationen müssen vor diesem gemeinsamen Hintergrund auf ihre Stimmigkeit überprüft werden. Durch diese Transparenz des Verstehens- und Interpretationsprozesses wird dieser auch notwendigerweise kritisierbar.

Diese Kriterien, die in ihrer Ausrichtung etwa dem konventionellen Kriterium der "internen Validität" entsprechen, fassen Guba und Lincoln unter dem Begriff der "Credibility" zusammen und entwickeln sechs methodische Zugänge zur Überprüfung dieser Glaubwürdigkeit (Guba und Lincoln 1989: 236 ff.).

In Korrespondenz zur "externen Validität" steht das Konzept der "Transferability" (Übertragbarkeit) nach Guba und Lincoln (1989: 241 f.) oder, nach Kvale (1991: 430 f.), dasjenige der "pragmatischen Validität". Die Validierung besteht dabei in der "Verifizierung im wörtlichen Sinne einer Wahrheitsprobe": Die Wirksamkeit der Erkenntnisgewinnung, die "Wahrheit der Schlussfolgerungen" muss sich in der Praxis, im konkreten Handeln beweisen. "Ein pragmatisches Konzept der Wahrheit ist dadurch charakterisiert, dass es uns dabei hilft, Handlungen in die Wege zu leiten, die zu den gewünschten Zielen führen" (Kvale 1991: 431).

12.2.3. Möglichkeiten und Grenzen qualitativer Evaluationsforschung

Qualitative, am interpretativen Paradigma orientierte Evaluationsforschung will in erster Linie dem Anspruch der Problemorientiertheit im Hinblick auf die Nützlichkeit genügen (vgl. Kapitel 13.3.). Ausgehend von praktischen Problemstellungen des Evaluationsbereiches sollen die Ergebnisse auf eine Optimierung des Vollzuges und eine höhere Wirksamkeit des Programms ausgerichtet sein. Die aus dieser Perspektive im Rahmen von Evaluationsstudien zu erarbeitenden Handlungsstrategien sollen die beteiligten und betroffenen Akteure auf einen politik- und sozialinnovatorischen Lernprozess hinführen.

Einzelfallbezogenheit, Problemorientierung und Kontextbezogenheit implizieren, dass die in die Evaluation miteinbezogenen Forschungssubjekte in ihrem natürlichen, alltäglichen Handlungskontext untersucht werden. Die Möglichkeit einer Generalisierbarkeit der Forschungserkenntnisse und -ergebnisse muss deshalb in jedem Fall argumentativ begründet werden. In diesem Sinn kann qualitative Evaluationsforschung dem positivistischen Anspruch auf schlüssige Verallgemeinerbarkeit nicht in der Art genügen, wie dies einzelne quantitative Ansätze (allerdings auch nur unter nicht sonderlich komplexen Bedingungen) zu tun vermögen. Der besondere Wert verstehensorientierter und interaktiver Evaluationsforschung liegt aber gerade in der Fähigkeit, Programme in einer hohen Komplexität ihrer Strukturen, Abläufe, Verfahren und Kontextsituationen sowie ihrer Entwicklungsdynamik zu erfassen und sie mit Hilfe der beteiligten Akteure zu interpretieren. In diesem Sinn legt dieser Evaluationsansatz sein Schwergewicht auf die problem- und einzelfallbezogene Nützlichkeit und nicht so sehr auf die Generalisierbarkeit der Evaluationsergebnisse.

12.2.4. Datenerhebung

12.2.4.1. Dokumentenanalyse

Nachdem der Rahmen des Evaluationsauftrages präzise definiert ist (Auftraggebende und deren Rolle im Evaluationsprozess, Definition des Evaluandums mit entsprechender Definition der Fragestellung, Auflistung der notwendigerweise einbezogenen Akteure), steht am Anfang jeder Evaluation eine seriöse Dokumentenanalyse. Sie verschafft das notwen-

dige Kontextwissen über die vorhandenen (und dem Forschungsteam zur Verfügung stehenden) Materialien, das es überhaupt erst erlaubt,

- die historischen Zusammenhänge der Fragestellung in ihren Grundzügen zu erfassen;
- das notwendige Fachwissen zum Gegenstandsbereich zu erlangen;
- als Forscher und Forscherin das eigene Vorverständnis gegenüber dem Gegenstandsbereich und seinen Kontexten zu entwickeln.

Das typische Merkmal der Dokumentenanalyse ist "die intensive, persönliche Auseinandersetzung mit dem Dokument, welches in seiner Einmaligkeit möglichst umfassend durchleuchtet und interpretiert wird" (Atteslander 1971: 67).

Da Dokumente als Datenmaterial nicht erhoben werden müssen, weil sie ja bereits fertig vorliegen, entstehen bei der Datenerhebung keine Schwierigkeiten, wohl können aber bei der Datenverfügbarkeit und -auswahl Probleme auftreten, die im Evaluationsbericht unbedingt transparent gemacht werden müssen:

- Aus politischen oder aus Gründen der Vertraulichkeit kann der Zugang zu wichtigen Dokumenten (oder bereits das Wissen um das Bestehen solcher Dokumente) verwehrt bleiben.
- Welche Dokumente in die Untersuchung einbezogen werden, ist ein subjektiver Entscheidungsakt der Evaluierenden.

Das Textmaterial kann etwa aus folgenden Dokumentenarten bestehen: gesetzliche Grundlagen; Weisungen; Strategie- und Konzeptpapiere; Sitzungs- und Tagungsprotokolle; Zeitungs- und Zeitschriftenberichte oder andere Publikationen. Auf der Basis der genau definierten Fragestellung muss vom Evaluator oder der Evaluatorin entschieden werden, welche Dokumente allein dazu dienen, das notwendige und möglichst umfassende Kontextwissen zu erlangen, das für eine schlüssige Interpretation der übrigen Daten dient, und welche Dokumente in einem theoriegeleiteten interpretativen oder quantitativen Analyse- und Auswertungsverfahren (quantitative und qualitative Inhaltsanalyse) zu erschliessen sind. (Eine ausführliche Darstellung der Theorie, der Methoden und der Praxis der Inhaltsanalyse findet sich in Merten 1983.)

12.2.4.2. Das problemzentrierte Interview

Das nicht-standardisierte qualitative Interview weist verschiedene Formen und Ausprägungen auf. Diese unterscheiden sich vor allem auf der Ebene der Strukturierung des Fragerasters.

Das *narrative Interview* z.B. zeichnet sich aus durch eine völlig offene Gesprächsführung; die Befragten sollen durch Erzählungen "Orientierungsmuster des Handelns" und eine "retrospektive Interpretation" des erzählten Handelns offenlegen (Lamnek 1989: 71). In diesem Sinn stellt das narrative Interview eine klar induktive Vorgehensweise dar; das Forschungsteam beginnt die Datenerhebung, ohne über - die Themenbereiche des Interviews strukturierende - wissenschaftliche Konzepte zu verfügen (Lamnek 1989: 74). Dies jedoch entspricht nicht einer idealen und realistischen Ausgangslage für eine Evaluation politischer Programme.

Eine zweite Form des qualitativen Interviews stellt das *fokussierte Interview* (nach Merton und Kendall 1984) dar. Diese Interviewform stützt sich zwar - wie beim problemzentrierten Interview auch - auf einen Interviewleitfaden. Dieser ist aber das Resultat einer detaillierten Theorie- und Hypothesenbildung durch die Forschungsequipe. Beim fokussierten Interview geht es um die Überprüfung von Hypothesen. Diese Interviewform ist deshalb stärker an einer positivistischen Forschungslogik orientiert (Lamnek 1989: 78).

Die dritte Form, auf welcher wir uns in Bezug auf die Evaluationsforschung konzentrieren, stellt das *problemzentrierte Interview* (Witzel 1985) dar. Seine wesentliche Eigenschaft besteht darin, dass es sich auf eine vordefinierte Problemstellung ausrichtet, die vom Interviewer in seinen wichtigsten Aspekten bereits im voraus erfasst (durch Literaturstudium und Dokumentenanalyse, durch Vorgespräche mit Experten im Untersuchungsfeld) und zu einem theoretischen Konzept verdichtet wurde. Diese Problemstellung wird im Verlauf des Interviews entlang eines Leitfadens im Rahmen eines kontinuierlichen Verstehensprozesses strukturiert, kommuniziert und interpretiert. Die partielle Standardisierung durch einen theoriegeleiteten Gesprächsleitfaden ermöglicht es, im Rahmen der Datenauswertung eine grössere Vergleichbarkeit der einzelnen Interviews zu erreichen.

Das problemzentrierte Interview basiert - trotz seiner themenbezogenen Strukturierung durch den Leitfaden - auf dem Erzählprinzip, d.h. die Auswahl, Reihenfolge und Sprachform dessen, worüber berichtet, was bewertet und interpretiert wird, bleibt den Befragten überlassen. In dieser

Interviewform bleibt auch mit einer gewissen Leitfadenbezogenheit viel Raum, in dem sich durch das frei assoziierende Erzählen wichtige Handlungszusammenhänge und zentrale kognitive Strukturen der Interviewpartner herauschälen. In diesem Sinn ist das problemzentrierte Interview nahe verwandt mit dem narrativen Interview.

Wie gross die Qualität und Dichte der Information ist, die im Rahmen eines Interviews ermittelt wird, hängt einerseits davon ab, wie gross die Fähigkeit des Befragers ist, den Interviewpartner zum Erzählen zu animieren und ihm den dazu notwendigen Raum zuzugestehen; von Bedeutung ist aber auch die Bereitschaft der Befragten, überhaupt erzählen zu wollen.

Kohli (1978) nennt als wichtigste Bedingung für den korrekten und ziel führenden Verlauf dieses Prozesses die Offenheit: Der Interviewte muss frei antworten können, ohne dabei durch vorgegebene Antwortkategorien eingeschränkt zu werden. Durch diese *Offenheit* kann erreicht werden, dass

- erkennbar wird, ob die Befragten verstanden haben, worum es geht und was der Interviewer erfahren möchte;
- die individuellen Sinngehalte und Bedeutungszuschreibungen, die kontextbezogenen kognitiven Strukturen und Interpretationen der Befragten und deren Sicht der Zusammenhänge innerhalb des erfragten Gegenstandsbereiches voll zum Ausdruck kommen können;
- den Befragten ein grösserer Aktivitätsraum zugestanden wird und diese dadurch die Möglichkeit zur eigenen Steuerung des Gesprächs haben. Offene Fragen erlauben den Befragten eigene Antwortformulierungen und lassen ihnen Raum zu längeren, komplexeren und assoziationsreichen Antworten;
- die Situation (raum-, zeit- und personenbezogen), in der das Interview stattfindet, von den Beteiligten thematisiert und in der nachfolgenden Interpretation durch das Evaluationsteam reflektierend miteinbezogen werden kann.

In partizipativen akteurbezogenen (stakeholder-based) Evaluationen sind problemzentrierte Interviews deshalb ein sehr adäquates Datenerhebungsinstrument, weil es im Rahmen des interpretativen Paradigmas und einem exploratorischen Untersuchungsdesign gerade darum geht, von den einzelnen Akteuren Situationsdeutungen, Handlungsmotive, Zweck-

Mittel-Relationen und entsprechende Selbstinterpretationen zu erfragen. Durch die "diskursive Verständigung über Interpretationen" sind mit offenen und teilstandardisierten Interviews die Möglichkeiten einer "empirischen Umsetzung handlungstheoretischer Konzeptionen" gegeben (Hopf 1991: 180).

Dem Forschungsteam dient zur Strukturierung des Gespräches ein an der Fragestellung orientierter *thematischer Leitfaden*. Dieser legt als eine Art Checkliste fest, welche Themenbereiche im Verlauf des Interviews anzusprechen sind. Die Einstiegsfragen in die einzelnen Themenbereiche werden im voraus formuliert und dienen im wesentlichen dazu, dass die Befragten - ohne sich unter einem intellektuellen Leistungsdruck zu fühlen - an ihren Alltagserfahrungen anknüpfen können. Die jeder Befragungssituation innewohnende "Künstlichkeit" (die beteiligten Personen sind einander fremd, ihre motivationale Beziehung ist asymmetrisch) kann dadurch etwas abgebaut und ein lockeres Interaktions- und Kommunikationsklima geschaffen werden. Bei Expertengesprächen wird von den Interviewten sehr häufig gewünscht oder verlangt, dass ihnen der Leitfaden vor dem Gesprächstermin schriftlich zugestellt wird. Diesem Begehren kann ohne methodische Bedenken entsprochen werden. Meistens hat diese Vorinformation bei Expertinnen und Experten auch zur Folge, dass diese sich inhaltlich auf die Befragung vorbereiten, was tendenziell zu umfassender und vollständiger Information führt.

Die Interviewer sollten sich in jedem Fall die sprachliche Ungebundenheit erhalten, die Fragen in einer der jeweils konkreten Interviewsituation und dem Interviewpartner angepassten Formulierung zu stellen. Nachfragen, i.S. von Verständnis- und Präzisierungsfragen, sollen auf die unmittelbar vorangegangenen Äusserungen der Befragten Bezug nehmen. Interviewer müssen sich in ihrer Gesprächsführung flexibel den Befragten, ihrem Erfahrungsbereich und ihren Sprachgewohnheiten anpassen. Die Reaktionsmöglichkeiten der Befragten sollten möglichst gross sein; diese müssen sich frei und ungezwungen äussern können. Müssen die Befragten mit offensichtlichen Widersprüchen und Ungereimtheiten konfrontiert werden, so sollte dies - im Hinblick auf die Erhaltung einer angenehmen Gesprächssituation - mit der nötigen Zurückhaltung erfolgen. Um authentische Äusserungen von hoher Verlässlichkeit zu erhalten, müssen Befragte über hinreichende Sach- und Sprachkompetenz verfügen, damit sie ihre Sicht der Realität, der Handlungen und der Kontexte hinreichend differenziert formulieren können.

In aller Regel werden Interviews auf *Tonband* aufgenommen. Dies ermöglicht einerseits, dass sich die Evaluatorinnen und Evaluatoren voll auf das Gespräch und dessen Ablauf konzentrieren können, ohne daneben durch das Erstellen von Notizen absorbiert zu sein. Andererseits gewährleisten Tonbandaufzeichnungen vollständige, authentische und von verschiedenen Personen bearbeitbare Gesprächsprotokolle, womit die Datenaufbereitung (Transkription) und Datenauswertung wesentlich erleichtert wird. Darüber hinaus kann eine Tonbandaufzeichnung später bei grösseren Differenzen in der Interpretation als Beweismittel dienen. In jedem Fall darf ein Gespräch nur mit dem ausdrücklichen Einverständnis der Befragten auf Tonband aufgenommen werden. Den Interviewten ist Vertraulichkeit bei der Verwendung der Tonbänder zu garantieren.

Die *Auswahl der Interviewpartner und -partnerinnen* richtet sich in erster Linie nach der expliziten Fragestellung der Evaluation - einschliesslich ihres thematischen, institutionellen und räumlichen Geltungsbereiches. Aus diesem Kontext ergeben sich die relevanten Akteure. Handelt es sich dabei um Einzelpersonen, so stellt sich die Frage der Auswahl nicht (allenfalls die Frage der Anzahl). Handelt es sich um eine Institution, so besteht möglicherweise ein Spielraum, welche der für den fraglichen Handlungsbereich repräsentativen Personen für eine Befragung auszuwählen sind³. Im Rahmen dieses Spielraumes ist es notwendig und legitim, jene Personen auszuwählen, die grundsätzliche Bereitschaft (wenn nicht gar Interesse) für ein solches Interview zeigen und die über die notwendige Sach- und Sprachkompetenz verfügen. Über die Kriterien und den Prozess zur Auswahl der Gesprächspartnerinnen und -partner muss Transparenz bestehen. Dabei ist durchaus zulässig, dass im Sinne eines iterativen Forschungsprozesses im Verlauf der Evaluation aufgrund gemachter Erfahrungen und des sich entwickelnden Erkenntnisstandes auch neue Personen einbezogen werden.

12.2.4.3. Gruppendiskussion

Bei der Gruppendiskussion als Methode der Datenerhebung geht es nicht um eine Erfragung von Einstellungen und Meinungen einzelner Personen, sondern um eine prozessorientierte - also situationsabhängige - Entwicklung von Meinungen und Interpretationen durch die Auseinandersetzung

³ Bei Politikevaluationen kommt es recht häufig vor, dass innerhalb einer Institution mehrere Akteure mit unterschiedlichen Aufgaben und Positionen zu einzelnen Massnahmen vorzufinden sind (vgl. Kapitel 4).

mit anderen Akteuren (Lamnek 1989: 136). Meinungen und Einstellungen sind sehr oft stark an gesellschaftlichen Zusammenhänge gebunden und können durch die Beteiligten demzufolge besser in sozialen Situationen - also in der Gruppe - offengelegt und erfasst werden. Durch Gruppendiskussionen ist es möglich, an "kollektive Einstellungen" heranzukommen (Mayring 1993: 54).

Zwar ist die Methode der Gruppendiskussion in der qualitativen Sozialforschung vorrangig aus dem Forschungsfeld der Meinungs- und Einstellungsforschung hervorgegangen (hier v.a. Pollock 1955); grösstenteils wurde mit Realgruppen gearbeitet (so etwa Niessen 1977). Dreher und Dreher schreiben dem Begriff "Diskussion" - in einer methodologisch weiter fassenden Art - die folgende Bedeutung zu: "Diskutieren wird methodologisch als valide Handlung erachtet, bei der durch den argumentativen Austausch von Begründungen und Bewertungen die Klärung eines Sachverhalts angestrebt wird, d.h. diskutieren ist exklusiv das Verfahren zur Erweiterung und Klärung von individuellen Sinnperspektiven in der Auffassung von Realität" (Dreher und Dreher 1991: 187). In dieser Anwendungsperspektive der Gruppendiskussion als "Basis edukativer Intervention" wird diese einerseits zur Informationsquelle für die Forschung, aber auch zum Lernprozess für die daran Beteiligten (Dreher und Dreher 1991: 187 f.).

Im Rahmen von Evaluationen kann die Methode der Gruppendiskussion eine wichtige und wertvolle *Ergänzung* zu den Einzelinterviews bilden. Sowohl mehrfach übereinstimmende als auch sich widersprechende oder kontrastierende Äusserungen und Interpretationen der einzelnen Befragten können der "Gruppe" der Akteure vorgelegt werden mit dem Ziel, im argumentativen Diskurs Differenzierungen, Ergänzungen und Begründungen der übereinstimmenden oder divergierenden Einzelaussagen und -interpretationen zu erreichen.

Innerhalb einer Evaluationsstudie steht die Methode der Gruppendiskussion nicht als alleiniges Datenerhebungsverfahren im Vordergrund. Die Gruppendiskussion eignet sich aber als Instrument für eine kommunikative und argumentative Validierung der in den Einzelgesprächen erhobenen Daten. So haben beispielsweise Bierter und Binder (1993) die in zahlreichen Einzelbefragungen gewonnenen Informationen sowie ihre Auswertung und Interpretation den beteiligten Akteuren im Rahmen von verschiedenen Workshops vorgelegt. Durch diese Auseinandersetzung mit den Forschungsergebnissen und den daraus ableitbaren künftigen Handlungsstrategien hatten die befragten Akteure die Möglichkeit, die

Interpretationen und Argumentationen der Forscher zu kritisieren und zu ergänzen. Die Gruppendiskussion wird damit zu einem zusätzlichen wichtigen Stein im Methoden-Mosaik, der es ermöglicht, den Untersuchungsgegenstand und sein Umfeld aus mehreren Blickwinkeln anzugehen und auszuleuchten. Diese Konzeption entspricht auch den Vorstellungen von Köckeis-Stangl (1980: 363) von einem Forschungsprozess im Sinne einer "mehrperspektivischen Triangulation", die ermöglicht, als Ergebnis kein einheitliches, sondern ein "kaleidoskopartiges" Bild des Untersuchungsgegenstandes zu erhalten. Damit gewinnt die Triangulation (vgl. Kapitel 11.3.) zu ihrem Wert als Validierungsstrategie zusätzlich die Bedeutung, dass durch sie in einer Untersuchung bewusst unterschiedliche Forschungsperspektiven miteinander verbunden und unterschiedliche Aspekte des Untersuchungsgegenstandes thematisiert werden (Flick 1991: 433).

12.2.5. Datenauswertung

Im Rahmen qualitativer Analyseverfahren geht es in der Phase der Datenauswertung darum, das in einer offenen und exploratorischen Form erhobene Datenmaterial sozialwissenschaftlich korrekt und praktikabel auszuwerten. Wie bereits im Abschnitt über das problemzentrierte Interview ausgeführt wurde, ist es angezeigt, Befragungen auf Tonband aufzunehmen. Das auszuwertende Datenmaterial aus Interviews und Gruppendiskussionen liegt also meistens in Form zahlreicher Tonbandprotokolle vor. Es stellt sich vorerst die Frage, wie die Fülle an Material hinsichtlich Form und Umfang verarbeitbar gemacht werden kann.

Das Verfahren der Textinterpretation muss so konzipiert sein, dass, von der ursprünglichen Fragestellung geleitet, das Datenmaterial in einem hermeneutischen Interpretationsprozess zusammengefasst, strukturiert und verstanden werden kann - und das mit einem Aufwand, der zeitlich wie personell den Ressourcen des Evaluationsauftrages entspricht. Die qualitative Sozialforschung hat in der jüngeren Vergangenheit verschiedene Auswertungsverfahren entwickelt (v.a. aus dem Bereich der Sozialpsychologie), die den Verstehensprozess von Sinnstrukturen mit hoher Transparenz und Präzision entwickeln und abbilden können. Es handelt sich dabei aber zumeist um Verfahren, die einen enormen Arbeitsaufwand erfordern und die im Falle von politikorientierten Evaluationen ein zu starkes Gewicht auf die detaillierte Analyse von Sprachcodes und deren

kontextbezogene Bedeutung legen. In diese Verfahrenskategorie gehören etwa die Ansätze der objektiven Hermeneutik (Oevermann 1979) und die Textinterpretationen nach Soeffner (1979). Wir werden uns deshalb im folgenden auf ein Auswertungsverfahren konzentrieren, das dem Anspruch der Machbarkeit genügt und das vorhandene Textmaterial entsprechend dem Erkenntnisinteresse politikorientierter Evaluation unter den erforderlichen Gütekriterien zu interpretieren vermag. In der Praxis hat sich in dieser Hinsicht das Verfahren der "qualitativen Inhaltsanalyse" nach Mayring (1983) bewährt. Diese Konzeption der Datenauswertung umfasst drei Grundcharakteristiken: Zusammenfassung, Explikation und Strukturierung. Mayring (1993: 86) fasst diese Merkmale wie folgt zusammen:

"Zusammenfassung: Ziel der Analyse ist es, das Material so zu reduzieren, dass die wesentlichen Inhalte erhalten bleiben, durch Abstraktion ein überschaubares Korpus zu schaffen, das immer noch Abbild des Grundmaterials ist.

Explikation: Ziel der Analyse ist es, zu einzelnen fraglichen Textteilen (Begriffen, Sätzen, ...) zusätzliches Material heranzutragen, das das Verständnis erweitert, das die Textstelle erläutert, erklärt, ausdeutet.

Strukturierung: Ziel der Analyse ist es, bestimmte Aspekte aus dem Material herauszufiltern, unter vorher festgelegten Ordnungskriterien einen Querschnitt durch das Material zu legen oder das Material aufgrund bestimmter Kriterien einzuschätzen." (Mayring 1993: 86).

Die Datenauswertung wird in mehreren Schritten durchgeführt. Als erster Schritt erfolgt die *Datenaufbereitung* in Form einer zweckmässigen *Transkription* der Tonbandprotokolle. Schon in diesem Verfahrensschritt geht es darum, die vorhandene Materialfülle zu reduzieren:

- Der abgehörte Text wird sprachlich bereinigt und geglättet: Übertragung vom Dialekt in die Schriftsprache; Satzbau- und Stilfehler werden behoben.
- Mehrfachnennungen, Wiederholungen werden zusammengefasst in ein einziges Zitat.
- Textstellen, die nichts mit der Fragestellung zu tun haben oder sehr nebensächlich sind, werden weggelassen.

In einem zweiten Schritt wird das gesamte transkribierte *Textmaterial* einem *beschreibenden Kategoriensystem* zugeordnet, d.h. einem Auswertungsraster, der aufgrund der Fragestellung entwickelt wird (Themenbe-

reiche, Evaluationskriterien). Dadurch erfolgt eine begründete Differenzierung der Fragestellung und eine Definition von Analysekatégorien, die letztlich bestimmen, welches Material für die Interpretation Bedeutung hat. Die Definition dieser Kategorien muss so genau und differenzierend sein, dass alle Textstellen, die im Transkript aus dem ersten Auswertungsschritt stehen, schlüssig einer Kategorie zugeordnet werden können.

Durch die Selektion und Bündelung von Aussagen (Paraphrasen, Zitate) wird das Textmaterial unter den vorgegebenen Auswertungskategorien zusammenfassend geordnet. Dadurch erfolgt gleichzeitig eine fallbezogene Reduktion und eine inhaltliche Strukturierung.

In einem nächsten Schritt geht es darum, die einzelfallbezogene Darstellung fallübergreifend zu *paraphrasieren* und zu *generalisieren*, um so zu einer Gesamtdarstellung zu gelangen. Wichtig ist, dass im Anschluss daran die betroffenen Akteure dazu befragt werden, ob sie mit dieser paraphrasierenden Interpretation einverstanden sind, ob sie sich in den entstandenen Aussagen wiedererkennen. Durch diese Kommunikation zwischen den Befragten und den Evaluationsteams kann die Gültigkeit der erfolgten Interpretation überprüft werden. Diese kommunikative Validierung bietet gleichzeitig die Gelegenheit zur Explikation, die ein integrierter Bestandteil der qualitativen Inhaltsanalyse ist: Durch direkte persönliche Nachfragen bei den Befragten können zu interpretationsbedürftigen Aussagen und Textstellen zusätzliche Erklärungen und Erläuterungen in die Analyse miteinbezogen werden.

Der Schritt der *kommunikativen Validierung* kann im Rahmen einer Gruppendiskussion (im Rahmen von strukturierten Workshops - siehe oben) erfolgen. Wenn die Evaluatorinnen und Evaluatoren ihre Interpretationen und ihre zugrundeliegenden Vorannahmen und theoretischen Bezüge gegenüber den befragten Personen offenlegen, findet eine argumentative Validierung statt: Durch die Präsentation nachvollziehbarer Interpretationen, Generalisierungen und Argumentationen werden diese kritisierbar. Eine weitere Form der kommunikativen und argumentativen Validierung ist der Diskurs innerhalb der Forschergemeinschaft sowie durch den breiteren Einbezug der Öffentlichkeit in die Interpretationsgemeinschaft (Kvale 1991: 430). Ein solcher Einbezug der Öffentlichkeit bedingt natürlich eine entsprechende Berichterstattung und Präsentation der Evaluationsergebnisse.

Partizipative und interaktive Validierung einer von den Forscherinnen und Forschern angestellten Interpretation von Daten ist nicht gleichzusetzen mit einer Art demokratischer "Beschlussfassung über den Schlussbe-

richt" oder über die Evaluation selbst. Die Evaluation basiert auf kommunikativ validierten Wirklichkeitsinterpretationen; sie ist damit indessen nicht zu verwechseln. Denn die Evaluation ist und bleibt eine mit nachvollziehbarem Design und kommunikativ validierten Daten erstellte Bewertung von Programm- und Massnahmenwirkungen, die einzig und allein die beauftragten Forschungsteams zu verantworten haben. Die partizipative Erhebungsmethode ist *ein* mögliches Mittel zur Erhebung gesellschaftlicher Wahrheit über Wirkungszusammenhänge, die dem konstruktivistischen Paradigma verpflichtet ist und daher eine relativ grosse Chance hat, von den Akteuren akzeptiert zu werden (auch wenn sie sich später gegebenenfalls als "falsch" herausstellen sollte) (vgl. Kapitel 13.3.).

12.3. QUANTITATIVE VERFAHREN

Dieser Abschnitt geht davon aus, dass eine Politikevaluation streng am rationalistischen Paradigma orientierte, hypothesenüberprüfende Vorgehensweise verfolgt. Dies impliziert einige Annahmen: Es bestehen bereits ausreichende Kenntnisse über die Strukturierung des Untersuchungsfeldes. Es liegt zudem ein deduktiv hergeleitetes Modell vor, das sich aus verschiedenen Hypothesen über die Wirkungszusammenhänge zusammengesetzt. Die Festlegung dieser Prämissen erfolgt aus didaktischen Gründen, um das Kapitel von Themen zu entlasten, die in diesem Band an anderen Stellen behandelt werden (vgl. Kapitel 12.1.).

Es kann in diesem Abschnitt nicht darum gehen, eine Einführung in die quantitativen Datenerhebungs- und Auswertungsverfahren zu geben, wie sie bereits in einer grossen Vielfalt in diversen Lehrbüchern zur sozialwissenschaftlichen Methodik vorliegen. Vielmehr soll es darum gehen, auf die spezifisch bei der Umsetzung dieser Verfahren in der Evaluationsforschung bestehenden Fragen einzugehen.

12.3.1. Datenerhebung

Die Erhebung quantitativen Datenmaterials kann einerseits so erfolgen, dass bereits bestehendes Material für eine Evaluation zugänglich gemacht wird. Anderenfalls sind die quantitativen Daten im Rahmen der Evalua-

tion eigens zu erheben. Im ersten Fall spricht man von einer Sekundärerhebung, im zweiten von einer Primärerhebung.

Für *Sekundärerhebungen* können unterschiedliche Datenquellen beigezogen werden. Dabei kann zwischen folgenden Typen unterschieden werden:

- prozessgenerierte Datenbestände,
- amtliche Statistiken und
- forschungsgenerierte Datenbestände.

Unter *prozessgenerierten Datenbeständen* versteht man Daten, die im politischen Prozess entstehen, die also beispielsweise im Rahmen des Vollzugs politisch-administrativer Massnahmen anfallen. Dazu gehören Informationen, welche die vollziehende Stelle für ihre eigenen Bedürfnisse bewusst und absichtlich herstellt (etwa im Bereich der Vollzugskontrolle). Daten können aber auch beiläufig als Nebenprodukt der Verwaltungstätigkeit anfallen. Prozessgenerierte Datenbestände weisen als besonderen Vorteil auf, dass sie - die Kooperation der Verwaltungsstelle vorausgesetzt und unter Vorbehalt der Datenschutzbestimmungen - oft für eine längere Zeitperiode verfügbar sind. Sie sind oft auch in einer nicht aggregierten Form zugänglich und zudem im Normalfall recht verlässlich. Ihre Nachteile liegen darin, dass sie nicht nach den Bedürfnissen einer Evaluation, sondern aufgrund der Erfordernisse der vollziehenden (oder der planenden) Verwaltung gesammelt und aufgezeichnet werden. Weiter ist damit zu rechnen, dass das Datenmaterial oft nicht für längere Zeitperioden anhand derselben Erfassungsregeln festgehalten wurde, sondern dass verschiedentlich Modifikationen auftreten, welche die Vergleichbarkeit der Daten in Frage stellen können. Dies gilt nicht nur für Vergleiche über die Zeit, sondern ebenso für Querschnittvergleiche. Oft sind die Differenzen der Erfassungsart der Daten zur selben Thematik zwischen verschiedenen öffentlichen Institutionen sehr gross. Teilweise lassen sich zwar derartige Mängel durch adäquate Umformungen umgehen, häufig ist dies jedoch nicht möglich. Weiter ist damit zu rechnen, dass Daten aus früheren Jahren oft nicht mehr existieren oder lediglich in einer Form vorliegen, die einen grossen Aufwand zur Aufarbeitung erfordert.

Unter Informationen aus *amtlichen Statistiken* versteht man Datenmaterialien, die durch spezielle statistische Dienste verwaltungsintern und in

Ausnahmefällen auch verwaltungsextern erstellt werden. Diese Daten sind üblicherweise frei zugänglich und im Normalfall auch seit Erfassungsbeginn verfügbar. Auch hier besteht zwar oft die Schwierigkeit, dass im Datenmaterial über die Zeit hinweg Strukturbrüche auftreten können, welche auf einen Wechsel der Erfassungsmodalitäten oder der Fragestellung zurückzuführen sind. Im Gegensatz zu den prozessgenerierten Daten sind aber hier oft genauere Informationen zu den Umstellungen erhältlich, die eine Rekonstruktion erleichtern können. Nachteilig auswirken kann sich bei Daten aus amtlichen Statistiken, dass die Informationen oft nur in aggregierter Form vorliegen und sich damit - je nach Fragestellung der Evaluation - als wenig nützlich erweisen. Weiter ist darauf hinzuweisen, dass die amtliche Statistik in der Schweiz, trotz der massiven Verbesserungen in den letzten Jahren, im internationalen Vergleich nach wie vor als rückständig bezeichnet werden muss. Die Botschaft des Bundesrates zum Beitritt der Schweiz zum Europäischen Wirtschaftsraum (EWR) sah vor, die Verwaltung im Bereich der amtlichen Statistik um 157 Stellen aufzustocken, um in diesem Bereich "europäisches Niveau" zu erreichen (BBl 1992 IV: 527-528). Zu verschiedensten Themen bestehen auch heute noch erhebliche Lücken. Vorteilhaft zeigen sich die Daten aus amtlichen Statistiken hinsichtlich der Form der Verfügbarkeit. Nicht nur ist der Zugang normalerweise leicht zu bewerkstelligen, auch stehen die Daten oft in maschinenlesbarer Form zur Verfügung, was - je nach Datenmenge - für die Evaluation grosse Erleichterungen bringen kann.

Als *forschungsgenerierte Datenbestände* bezeichnet man Datenmaterial, das im Zusammenhang mit einem Forschungs- oder Evaluationsprojekt entstanden ist. Diese Datenbestände sind somit auf eine spezifische Fragestellung zugeschnitten. Es kann sich jedoch als zweckmässig erweisen, dieses Datenmaterial im Rahmen einer Evaluation einer sekundären Nutzung zuzuführen. Dafür müssen die Daten jedoch zumindest Teile der aktuellen evaluativen Fragestellung abdecken. Meistens ist diese Bedingung aber nicht vollumfänglich erfüllt, so dass die forschungsgenerierten Datenbestände mit weiteren Informationen ergänzt werden müssen. Es kann sich aber als sehr nützlich erweisen, auf diese Daten zurückzugreifen, da damit ressourcenintensive Erhebungen entfallen können. Problematisch ist bei dieser Datenquelle aber - leider - sehr oft der Zugang. Viele Primärforscherinnen und -forscher geben eigens erhobenes Datenmaterial nur sehr ungern an Dritte weiter. Oder sie tun dies erst zu einem Zeitpunkt, in dem das Material deutlich an Aktualität verloren hat. Dies-

bezüglich interessant sind die Datenvermittlungsdienste, welche der "Schweizerische Informations- und Daten-Archivdienst für die Sozialwissenschaften" (SIDOS) anbietet. Es ist zu hoffen, dass dadurch die Sekundärnutzung von forschungsgenerierten Datenbeständen (auch in der Schweiz) zunehmend erleichtert wird (vgl. dazu SIDOS 1994).

Sehr oft erweisen sich die zur Verfügung stehenden Daten aus den obengenannten Quellen nicht als ausreichend, um die Evaluation zu bearbeiten, so dass sich eine *Primärerhebung* aufdrängt. Dazu stellt die sozialwissenschaftliche Methodenlehre verschiedenste Instrumente zur Verfügung. Für Zwecke der Evaluationsforschung bieten sich dafür insbesondere folgende Erhebungsverfahren an:

- standardisierte Befragung
- Beobachtung / Messung
- quantitative Inhaltsanalyse
- Indizes und Skalen

Während bei den beiden erstgenannten Techniken eine eigentliche Datenproduktion erfolgt, zielen die beiden anderen darauf ab, bereits bestehende Quellen derart aufzuarbeiten, dass die Informationen einer quantitativen Datenauswertung zugeführt werden können. Die obige Aufzählung ist in keiner Weise als vollständig zu betrachten. Für weitere Erhebungstechniken, die sich ebenfalls für Evaluationen eignen können, wird auf die einschlägige Literatur verwiesen (bspw. Atteslander 1993; Friedrichs 1984; Kromrey 1994; Schnell, Hill und Esser 1993).

Die *standardisierte Befragung* ist wohl das bekannteste Verfahren zur sozialwissenschaftlichen Datenerhebung und auch in der Evaluationsforschung weit verbreitet (Weiss 1975). Standardisiert meint, dass allen Befragten die gleichen Fragen vorgelegt werden. Um die Befragungen vergleichbar zu machen, versucht man dabei die Bedingungen möglichst konstant zu halten. Zur Durchführung von standardisierten Befragungen stehen verschiedene Formen zur Verfügung:

- persönliche Befragungen
- telefonische Befragungen
- schriftliche Befragungen

Bei einer persönlichen Befragung werden die Daten in direkter Interaktion zwischen Interviewer und Befragten erhoben. Der Interviewte sucht dabei üblicherweise den Befragten zu Hause auf. Es kommt jedoch auch vor, dass solche Interviews an anderen Orten - also etwa auf der Strasse - stattfinden. Beim telefonischen Interview wird die Person am Telefon befragt. Dabei kommt in letzter Zeit häufig ein "computer assisted telephone interviewing"-System, abgekürzt CATI, zum Einsatz. Solche Systeme erleichtern es den Interviewern, sich im Befragungsprozess zurechtzufinden, und ermöglichen den Einsatz komplexerer Fragebogenkonstruktionen. Ein weiterer Vorzug der CATI-Technik besteht darin, dass der Interviewer im Laufe des Telefongesprächs die Antworten in maschinenlesbarer Form erfassen kann, womit ein sonst notwendiger Arbeitsschritt eingespart werden kann. Bei einer schriftlichen Befragung werden die Fragebogen den Befragten - normalerweise per Post - zugestellt. Diese füllen die Fragebogen ohne die Unterstützung eines Interviewers selbst aus und retournieren den ausgefüllten Fragebogen. Ferner können diese Befragungsformen auch kombiniert werden, etwa so, dass im Anschluss an eine persönliche Befragung die befragte Person darum gebeten wird, einen ergänzenden schriftlichen Fragebogen auszufüllen. Es empfiehlt sich, die Auswahl der Befragungsform überlegt zu treffen, da mit den verschiedenen Formen unterschiedlichste Vor- und Nachteile verbunden sind (für genauere Erläuterungen siehe Weiss 1975; Holm 1991; GAO 1991b: 11-17).

Die Befragungstechnik ist nicht nur sehr weit verbreitet, sondern es wurden auch eine Vielzahl verschiedener Regeln geschaffen, um Befragungen valider und reliabler zu gestalten (siehe etwa die Hinweise in Atteslander 1993: 126-200; Schnell et al. 1993: 328-389; Holm 1991). Diese Regeln sollten beachtet und vor allem auch konsequent umgesetzt werden.

Beobachtung und Messung werden in der Evaluationsforschung relativ häufig eingesetzt, wobei deren Einsatz oft zu wenig reflektiert wird. Aufgrund des interdisziplinären (oder zumindest multidisziplinären) Charakters der Evaluationsforschung befassen sich Evaluationsstudien oft mit Untersuchungsgegenständen, welche im strengen Sinn nicht oder nur teilweise sozialwissenschaftlicher Art sind. Vor allem Fragestellungen aus den Gegenstandsbereichen der Naturwissenschaften sind häufig in Evaluationen anzutreffen. Man denke dabei etwa an die verschiedenen Projekte zur Abklärung der Wirksamkeit staatlicher Umweltschutzmassnahmen (so Brandner et al. 1995; Balthasar und Knöpfel 1994; Knoepfel und

Zimmermann 1987; Widmer 1991). Aber auch die Ingenieurwissenschaften sind - etwa im Bereich der Energiepolitik - stark vertreten (siehe etwa Linder, Looser, Trombitas und Wittwer 1990, Aebischer et al. 1993). Das gleiche gilt für die Medizin, die bei Wirksamkeitsüberprüfungen von gesundheitspolitischen Massnahmen wichtige Beiträge einbrachte (siehe bspw. Dubois-Arber, Jeannin, Meystre-Agustoni, Gruet und Paccaud 1993; Bahy 1986). Durch die Art der in der Evaluationsforschung zu bearbeitenden Fragestellungen drängen sich daher oft Verfahren wie Beobachtung oder Messung auf. Auch hier gilt, dass es sich um Verfahrenstechniken handelt, die eine lange Entwicklung hinter sich haben, in deren Verlauf das Instrumentarium verfeinert wurde. Dabei sind eine Vielzahl von Verfahrensregeln entstanden, die berücksichtigt werden sollten. Während die Beobachtung in einigen sozialwissenschaftlichen Disziplinen eine recht weite Verbreitung gefunden hat, sind eigentliche Messtechniken doch eher im Bereich der exakten Wissenschaften anzutreffen. Reichen dafür die Kenntnisse des Evaluators selbst nicht aus, ist es zweckmässig, für Datenerhebungen dieser Art externe Expertise einzuholen. Im Idealfall wird ein multidisziplinär zusammengesetztes Projektteam gebildet.

Die *quantitative Inhaltsanalyse* erlaubt eine Umsetzung verbaler Daten in quantitative Angaben, welche eine quantitative Auswertung erlauben. Als Rohdaten dienen jegliche Quellen in Textform, wie etwa Dokumente, Verwaltungsakten, Zeitungsartikel, Transkripte von Reden, Interviews, Diskussionen usw. Im Gegensatz zur qualitativen Inhaltsanalyse (vgl. Abschnitt 12.2.) geht es bei der quantitativen Inhaltsanalyse also darum, aus qualitativen Informationen quantitative Angaben zu generieren. Dies erfolgt zum Beispiel in der Form, dass ein gegebener Text entlang eines bestehenden theoretischen Codierschemas, welches die relevanten Kategorien umfasst, durchgearbeitet wird. Dabei werden die den Kategorien zugehörigen Begriffe, Sätze, Sequenzen oder andere Textteile (bewertet und) ausgezählt, womit numerische Angaben entstehen, die einer Datenauswertung zugeführt werden können. Das Ziel besteht darin, eine quantitative Erfassung eines Textes zu erreichen, welche intersubjektiv zu identischen Ergebnissen führt. In diesem Prozess gehen natürlich im Text enthaltene Informationen, wie etwa die Position einer bestimmten Aussage im Text, also der Kontext einer Aussage, verloren. In dieser Art lassen sich sowohl kürzere wie auch längere, zahlreiche wie auch nur wenige oder gar ein einziger Text analysieren. Für diese Zwecke existieren bereits verschiedene Computerprogramme, welche die Arbeit - je nach Fragestellung und Analyseform - massgeblich erleichtern können (siehe

dazu Weber 1990). Zur quantitativen Inhaltsanalyse besteht auch bereits eine ansehnliche Literatur, die vorteilhafterweise vor dem Einsatz dieser Technik zu konsultieren ist (Krippendorff 1980; Merten 1983; Lisch und Kriz 1978).

Indizes und Skalen erlauben es dem Evaluator, bestehendes quantitatives Datenmaterial zu inhaltlich miteinander verbundenen Indikatoren in einer Dimension zu erfassen. Das heisst, es werden verschiedene quantitative Indikatoren in einen Index oder eine Skala überführt, welche(r) die Informationen der Rohdaten möglichst gut repräsentieren soll. Der Unterschied zwischen Indizes und Skalen sei im folgenden kurz erläutert. Ein *Index* versucht, Indikatoren zu verschiedenen Teilaspekten zu einem Indikator (Index genannt) zusammenzufassen. Es liegen also empirische Angaben zu verschiedenen Teilkonzepten eines Konstrukts vor. Der resultierende Index soll das Gesamtkonstrukt möglichst gut darstellen. Die verschiedenen Einzelindikatoren werden dabei zumeist additiv, seltener multiplikativ miteinander verbunden. Das in der Schweiz wohl prominenteste Beispiel stellt der sogenannte Landesindex der Konsumentenpreise (Koch 1990: 283-297) dar, der aufgrund der Preisangaben verschiedenster Konsumprodukte (aus dem sogenannten Warenkorb) aufgrund einer festgelegten Gewichtung errechnet wird. Eine *Skala* hingegen misst ein Konstrukt zum Beispiel in einem Fragebogen - anhand von einzelnen Fragen (Items genannt), die jeweils das *gleiche* Konstrukt in unterschiedlich starken Ausprägungen erfassen. Zur Erstellung von Skalen stehen eine grosse Anzahl von unterschiedlichen Skalierungsverfahren zur Verfügung (Mokken-, Guttman-, Likert-Skala etc.; vgl. Bortz 1984: 149-159; Schnell et al. 1993: 193-221; Friedrichs 1984: 172-188). Für ähnliche Fragestellungen bestehen - neben Indizes und Skalen - weitere Verfahren wie die Cluster-Analyse oder die Faktorenanalyse, die auf komplexeren statistischen Modellen beruhen (Aldenderfer und Blashfield 1984; Backhaus, Erichson, Plinke und Weiber 1994; Bortz 1989: 684-705 bzw. 613-683; Kim und Mueller 1979).

Mit den hier kurz beleuchteten Erhebungsverfahren lassen sich quantitative Daten unterschiedlicher Art erheben. Von Bedeutung sind hier besonders die folgenden Eigenschaften:

- *Messniveau*: Das Messniveau gibt die Qualität der Messskala an, mit der die Indikatoren gemessen werden. Es wird (grob) unterschieden zwischen Nominal-, Ordinal- und Kardinalskalen. Bei nominalskalierten Daten ist lediglich die Unterscheidung der verschiedenen Ausprä-

gungen möglich (wie der Eigenname, die Parteizugehörigkeit, der Wohnort oder die Telefonnummer einer Person). Ordinale Skalen zeichnen sich dadurch aus, dass zusätzlich die verschiedenen Werte in eine Reihenfolge gebracht werden können (wie Schulnoten oder die soziale Schichtzugehörigkeit). Die Klassen der nominalen und ordinalen Messniveaus werden auch als kategorial bezeichnet. Kardinalskalierte Daten dagegen erlauben zudem, die Differenzen zwischen den verschiedenen Ausprägungen miteinander zu vergleichen (wie Alter, Einkommen, Temperatur). Diese Skalentypologie kann weiter differenziert werden. So umfasst die Kategorie der kardinalen Skalen sowohl Intervallskalen (wie Temperatur in °C), wo kein absoluter Nullpunkt besteht, wie auch Ratio-Skalen, welche über einen absoluten Nullpunkt verfügen (wie Einkommen, Temperatur in Grad Kelvin). Die Messniveaus der erfassten Indikatoren sind deshalb wichtig, weil sich die Interpretationsmöglichkeiten je nach Skalentyp stark unterscheiden. Weiter sind sie auch bei der Auswahl der adäquaten Analyseverfahren von grosser Bedeutung (vgl. Abschnitt 12.2.). Das Messniveau erhobener Daten kann zwar, beispielsweise mit Kategorisierungen, verändert werden, jedoch nur in einer Richtung. Es ist möglich, aus einer Kardinalskala eine Ordinal- oder Nominalskala zu generieren (was jedoch stets mit einem Informationsverlust verbunden ist). Es besteht jedoch ohne zusätzliche Informationen keine Möglichkeit, aus einer Nominalskala nachträglich eine Ordinal- oder Kardinalskala zu produzieren, da die im empirischen Material enthaltenen Informationen dazu nicht ausreichen.

- **Erhebungsebene:** Vor der Datenerhebung ist die Erhebungseinheit zu definieren. Es muss festgelegt werden, worauf sich die Variablen beziehen sollen. Es wird dabei grob in Individualdaten und Aggregatdaten unterschieden. Bei Individualdaten bildet ein Individuum (Person, Entscheidung, Parzelle etc.) die Erhebungseinheit (Mikroebene), bei Aggregatdaten dagegen eine Gruppe von Individuen, eine Institution (Mesoebene), eine Gemeinde oder ein Staat (Makroebene). Auch dieser Entscheid ist bereits in der Phase der Datenerhebung zu treffen und bestimmt massgeblich die Interpretationsmöglichkeiten. Aufgrund der Analyse von Aggregatdaten sind keine Interpretationen auf individueller Ebene möglich (sog. ökologischer Fehlschluss, vgl. dazu Robinson 1950; Erbring 1990). Es ist hingegen möglich, Individualdaten auf der Meso- oder Makro-Ebene zu aggregieren. Anders ausge-

drückt bedeutet dies, dass die Erhebungseinheit sehr wohl kleiner aber niemals grösser als die Analyse- oder die Aussageeinheit sein kann.

- **Erhebungsumfang:** Zu Beginn einer Datenerhebung ist zu entscheiden, ob Datenmaterial für alle möglichen Untersuchungseinheiten oder nur für eine Auswahl (eine sog. Stichprobe; vgl. Kapitel 11.1.) erhoben werden soll. Insbesondere wenn eine Grundgesamtheit mit einer grösseren Anzahl von Elementen vorliegt, wird in der quantitativen Sozialforschung sehr häufig mit Stichproben gearbeitet (siehe Kaplitza 1991; GAO 1992a; Bortz 1984: 239-362). Es bestehen verschiedene Verfahren zur Ziehung von Stichproben. Weiter muss festgelegt werden, wie gross die Stichprobe sein soll. Auch diese Entscheide haben einen Einfluss auf die Analyse des Datenmaterials und auf deren Interpretation. So sind bei Interpretationen von Stichprobendaten immer auch die statistischen Fehler im Auge zu behalten. Zudem erhalten die in der Analyse resultierenden Signifikanzen (zur Erläuterung vgl. Abschnitt 12.3.2.) von Analyseergebnissen in diesen Fällen eine weitreichende Bedeutung. In seltenen Fällen kommt es vor, dass nach der Durchführung einer Vollerhebung eine Stichprobenziehung erfolgt (z.B. Klöti und Risi 1991: 267-283).
- **Erhebungszeitraum:** Je nach gewählttem Untersuchungsdesign (vgl. dazu Kapitel 11) sind Daten zu nur einem Zeitpunkt oder zu verschiedenen Zeitpunkten (oder Zeiträumen) zu erheben. Der Zeitpunkt bzw. der Zeitraum sind präzise zu definieren. Soll eine zeitkontinuierliche Betrachtung erfolgen, ist zu entscheiden, in welchem Zeitintervall Daten erhoben werden sollen (Stunden-, Tages-, Wochen-, Monats-, Quartals- oder Jahresdaten). Die Auswahl des Erhebungszeitraums soll an die Fragestellung angepasst und adäquat zum gewählten Untersuchungsdesign erfolgen. Weiter ist darauf zu achten, dass kurzfristige Veränderungen (innert Tagen, Wochen oder Monaten) nur in unzureichender Form mit längerfristig angelegten Daten (Jahresdaten) erfasst werden können. In kurzfristigen Intervallen erhobene Daten können auf längere Intervalle aggregiert werden (z.B. Widmer 1991: 93-101).

Auch die Frage, welche Erhebungsart(en) im spezifischen Evaluationsprojekt zu wählen ist (sind), kann nur aufgrund der konkreten Fragestellung, der gegebenen Rahmenbedingungen und der empirischen Datenlage

beurteilt werden. Trotzdem lassen sich aber einige grundlegende Leitlinien formulieren.

- Bestehendes Datenmaterial, das zur Klärung der Fragestellung nützlich sein und mit vergleichsweise kleinem Aufwand erhoben werden kann, sollte berücksichtigt werden.
- Es sollte immer darauf geachtet werden, dass für die Auswertung des erhobenen Datenmaterials ausreichend Mittel zur Verfügung stehen. Es macht keinen Sinn, "Datenberge" zu produzieren, für deren Auswertung die Ressourcen nicht ausreichen.
- Datenerhebungen sollten immer in Abstimmung mit der Fragestellung erfolgen. Es hat keinen Sinn, Informationen zu erheben, die keinen Beitrag zur Analyse leisten können.
- Sekundärerhebungen sind im Regelfall günstiger und schneller durchführbar als Primärerhebungen.
- Primärerhebungen haben den Vorteil, dass sich die Datenerhebung exakt auf die Bedürfnisse der Evaluation ausrichten lassen. Bei Sekundärerhebungen ist dies nicht beeinflussbar.
- Auch bei Sekundärerhebungen ist der Entstehungszusammenhang des Materials bei der Datenauswertung und Interpretation zu berücksichtigen.

Erhobenes Datenmaterial ist hinsichtlich seiner Qualität mittels Gütekriterien (wie Validität und Reliabilität, vgl. dazu Schnell et al. 1993: 156-176; Cedro und Gualtierotti 1996; mit einem breiteren Ansatz Patton 1984) zu beurteilen. Dies gilt sowohl für Primär- wie auch für Sekundärerhebungen. Ist die Datenqualität aufgrund dieser Überprüfung nicht in ausreichender Weise gegeben, ist dies bei Analyse und Interpretation zu berücksichtigen. Im Extremfall muss auf eine Auswertung der Daten gänzlich verzichtet und auf anderes Datenmaterial ausgewichen werden. Weiter ist eine systematische Überprüfung des Datenmaterials durchzuführen (Plausibilität, Ausschliessen fehlerhafter Codes etc.), um zu vermeiden, dass die Evaluation fehlerhafte Angaben enthält. Zweckmässigerweise erfolgt diese Überprüfung auch nach der Erfassung der Daten in maschinenlesbarer Form, weil diese Übertragung ebenfalls eine Fehlerquelle darstellen kann. Sind die Daten in einem EDV-System erfasst, kann die Datenüberprüfung mit EDV-Unterstützung erfolgen.

12.3.2. Datenauswertung

Es ist nicht möglich, in diesem Kapitel einen nur im Ansatz vollständigen Überblick über die diversen quantitativen Auswertungsverfahren zu bieten (vgl. dafür die einschlägige Einführungsliteratur, wie Bortz 1989; Wonnacott und Wonnacott 1990; Kriz 1983; GAO 1992b und die darin enthaltenen Literaturverweise). Zur Wahl der adäquaten Datenanalysetechnik hat sich eine Broschüre als äusserst hilfreich erwiesen, welche mittels eines Entscheidungsbaums zum passenden Verfahren führt (Andrews, Klem, Davidson, O'Malley und Rodgers 1981). Im folgenden soll es darum gehen, einige häufig eingesetzte Analysetechniken exemplarisch herauszugreifen und gewisse mit diesen Verfahren in Zusammenhang stehende Aspekte aufzugreifen. Weiterhin wird versucht, jeweils beispielhaft auf praktische Umsetzungen dieser Verfahren in der (schweizerischen) Evaluationsforschung hinzuweisen. Zur Strukturierung des Kapitels verwenden wir folgende drei Dimensionen in der angeführten Hierarchie:

- *Vergleichsdimension*, also Querschnitt-, Längsschnittanalysen und kombinierte Verfahren,
- *Dimensionalität der Analyse*, also univariate, bivariate und multivariate Analysen und
- *Messniveau des Datenmaterials*, also Verfahrenstechniken zu nominalem, ordinalem und kardinalen Datenmaterial⁴.

Diese Dimensionen sind in keiner Weise als erschöpfend zu betrachten, genügen aber für die vorgesehene kurze Übersicht. Die nachfolgende Tabelle zeigt mögliche Verfahren bei Querschnittanalysen:

⁴ Hierzu drängt sich noch eine Zusatzbemerkung auf. Ist eine Analyse mit Daten vorgesehen, welche über unterschiedliche Messniveaus verfügen (bspw. ordinale und kardinale Variablen), ist für die Wahl des Analyseverfahrens im Regelfall das "niedrigste" Messniveau (in unserem Beispiel die Ordinalskala) ausschlaggebend.

Abbildung 8 Dimensionalität und Messniveaus ausgewählter Verfahren in Querschnittanalysen

	<i>nominal</i>	<i>ordinal</i>	<i>kardinal</i>
<i>univariat</i>	Häufigkeitsauszählung (mit Prozentuierung), Modalwert		Masse der zentralen Tendenz, Dispersionsmasse
<i>bivariat</i>	Kreuztabellen- analyse, ϕ (Phi) bzw. Cramers V und λ (Lambda) u.a., χ^2 (Chi-Quadrat)	Kreuztabellen- analyse, Yule's Q bzw. Gamma (γ), Kendalls tau-b, Stuarts tau-c, Somers d, χ^2 (Chi-Quadrat)	Korrelations- analyse (r), bivariate Regressions- analyse (R^2)
<i>multivariat</i>	Log-lineare Modelle, Latente Klassenanalyse (LCA), Lineare Strukturgleichungsmodelle (LISREL, Version 8)		multiple Regression, Kausalanalyse, Lineare Struktur- gleichungsmodelle (LISREL)

Wie aus der Tabelle ersichtlich wird, grenzen die Dimensionalität der Analyse und die Messniveaus der erhobenen Daten die Auswahl an möglichen Analyseverfahren bereits stark ein. Dies sei an der *univariaten Analyse* mittels Massen der zentralen Tendenz (wie arithmetisches Mittel, Median oder Modalwert) erläutert. Wie aus der Tabelle hervorgeht, ist der Modalwert bei ordinalen oder nominalen Daten das einzige hier zulässige Mass. Die Berechnung des arithmetischen Mittels ist hier also zum Beispiel nicht zulässig. Verwendet man zur Analyse von Häufigkeitsauszählungen prozentuierte Angaben, ist darauf zu achten, dass die zugrunde liegenden Häufigkeiten ausreichend gross sind. Ein absolutes Muss besteht dabei in der Angabe der Fallzahl (N resp. n), der sog. Prozentuierungsbasis. Zur Analyse der Streuung empirischer Verteilungen auf kardinaler Ebene bietet sich eine Vielzahl verschiedener Masse an. Genannt sind hier exemplarisch die Varianz (s^2), die Standardabweichung (s), verschiedene Quantile (Quartile, Zentile etc.), die Streubreite ("range") und der Variationskoeffizient (CV). Diese Masse haben jeweils

ihre spezifische Charakteristik. So dient s üblicherweise dazu, zu beurteilen, wie gross die Streuung im Verhältnis zu den absoluten Werten ist. Sie hat gegenüber anderen Massen den Vorzug, dass sie dieselbe Einheit aufweist wie die absoluten Werte. Möchte man dagegen zwei empirische Verteilungen bezüglich ihrer Dispersion miteinander vergleichen, ist dies mit s nur möglich, wenn die Mittelwerte der beiden Verteilungen gleich gross sind. In den übrigen Fällen ist der Vergleich verzerrt. Deshalb wird für derartige Vergleiche üblicherweise der CV verwendet. Dieser Koeffizient ist sog. mittelwertbereinigt, d.h., das Mass ist um die Mittelwerte der Verteilungen korrigiert. Die univariaten Verfahren können im Kontext der Evaluationsforschung einerseits als eigenständige Analyseverfahren eingesetzt werden (für Beispiele dazu siehe etwa Gerheuser und Schmid 1993: 12-18; Schulz, Muggli und Hübschle 1993: 73-119; Balthasar und Knöpfel 1994: 104-118). Andererseits ist es sehr empfehlenswert, die erhobenen Daten vor einer bi- oder multivariaten Analyse einer ausführlichen univariaten Analyse zu unterziehen. Erstens können damit Erkenntnisse gewonnen werden, die in der mehrdimensionalen Analyse verborgen bleiben, andererseits bildet die univariate Analyse eine notwendige Vorarbeit für weitergehende Analyseschritte (Beispiel: Überprüfung der Verteilungsannahmen).

Mit den *bi- und multivariaten Verfahren* werden Zusammenhänge zwischen verschiedenen Variablen untersucht. Es sei an dieser Stelle deutlich darauf hingewiesen, dass (alleine) mit empirischen Datenauswertungen nicht auf *kausale* Zusammenhänge geschlossen werden kann. Die empirische Analyse ist lediglich in der Lage, eine theoretisch hergeleitete Kausalität zu überprüfen. Die *Kreuztabellenanalyse* ist im Prinzip eine zweidimensionale Häufigkeitsauszählung. Es wird überprüft, ob zwischen kategorialen (also nominalen bzw. ordinalen) Variablen ein empirischer Zusammenhang vorliegt. In der vorangehenden Abbildung 8 finden sich die für das jeweilige Messniveau adäquaten Assoziationskoeffizienten. Diese erlauben Aussagen über die Stärke eines Zusammenhangs. Zur Feststellung der Signifikanz eines Zusammenhangs wird üblicherweise ein sogenannter χ^2 -Test durchgeführt. Dieser lässt die Beurteilung zu, ob ein bestimmter Zusammenhang auf Zufall beruht oder ob eine strukturelle Regelmässigkeit vorliegt. Aussagen dieser Art sind nie "sicher", sondern immer einer bestimmten Irrtumswahrscheinlichkeit ausgesetzt. Die Grösse dieser Fehlerwahrscheinlichkeit, welche man in Kauf zu nehmen bereit ist, wird mit dem sog. Signifikanzniveau festgelegt. Im sozialwissenschaftlichen Bereich wird üblicherweise von fünf- oder einprozentigen

Irrtumswahrscheinlichkeiten ausgegangen (in Kurzschreibweise $p < .01$ oder $p < .05$). Handelt sich bei dem untersuchten Datenmaterial um eine Zufallsstichprobe (oder annähernd um eine solche) und erweist sich ein Zusammenhang als signifikant, kann das Analyseergebnis mit der festgelegten Irrtumswahrscheinlichkeit auch auf die Grundgesamtheit übertragen werden. Analoge Signifikanztests - wie das χ^2 -Mass bei Kreuztabellen - bestehen auch bei den anderen Verfahren zur Überprüfung von Zusammenhängen (bei der Korrelationsanalyse wird üblicherweise eine t-Test-Statistik berechnet).

Die *Korrelationsanalyse* verfolgt im Prinzip den gleichen Zweck wie die Kreuztabellenanalyse. Auch hier geht es um die Überprüfung bivariater Zusammenhänge, allerdings bei kardinalen Daten. Die sog. Produkt-Moment-Korrelation nach Pearson ist das am weitesten häufigsten verwendete Verfahren in der Korrelationsanalyse. Es resultiert ein Korrelationskoeffizient, der Pearsons r genannt wird. Dieser Koeffizient lässt Aussagen über die Stärke des empirischen Zusammenhangs zu. Die Korrelationsanalyse geht von der Annahme aus, dass der Zusammenhang eine lineare Form aufweist. Das bedeutet, dass sich die Werte der beiden untersuchten Variablen proportional zueinander verhalten. Dies kann mittels eines Streudiagramms optisch überprüft werden. Der resultierende Punkteschwarm, ein Punkt steht für eine Beobachtung, sollte annäherungsweise linear verlaufen. Auf die Überprüfung der Linearitätsannahme sollte keinesfalls verzichtet werden. Nicht-lineare Zusammenhänge sind im Bereich der Evaluationsforschung relativ häufig anzutreffen und die Auswirkungen eines Verstosses gegen diese Annahme verzerrt die Resultate massiv.

Quasi eine Erweiterung der Korrelationsanalyse bildet die *bivariate Regressionsanalyse*. Mit Hilfe dieser Methode versucht man eine sog. Regressionsgerade zu schätzen, welche den Punkteschwarm im Streudiagramm möglichst weitgehend repräsentiert. Aus einer Regressionsanalyse resultiert ein sogenannter Determinationskoeffizient R^2 , welcher Aussagen darüber erlaubt, zu welchem Anteil die Varianz in einer Variable (als abhängige (oder endogene) Variable (y) bezeichnet) durch die andere Variable (als unabhängige (oder exogene) Variable (x) bezeichnet) "erklären" kann. Es handelt sich auch hierbei nicht um eine kausale, sondern lediglich um eine statistische Erklärung. Weiter wird auch die Steigung (mit b bezeichnet) der Geraden geschätzt. Anhand des Schätzwertes b lassen sich Folgerungen formulieren, wie beispielsweise: "Wenn x um eine Einheit ansteigt, verändert sich y um den Wert b ." Bivariate Regressionsanalysen lassen sich auch durchführen, wenn eine nicht-lineare Be-

ziehung vorliegt, oder wenn andere Verstösse gegen die Grundannahmen einer Regressionsschätzung vorliegen. Das konventionelle Schätzverfahren (Methode der kleinsten Quadrate ("ordinary-least-squares", abgekürzt OLS) ist dazu jedoch weniger gut geeignet als andere, komplexere Schätzverfahren (siehe dazu Gujarati 1988; Stewart 1991; für Ambitionierte: Johnston 1972). Es bestehen auch Weiterentwicklungen der Regressionsmethodik, die überdies die Auswertung nicht-kardinaler Daten ermöglichen (so etwa die Logit-Analyse, siehe Urban 1993; Demaris 1992). Ein Beispiel für die Anwendung höher entwickelter Regressionsverfahren findet sich in einer Evaluationsstudie zu den Auswirkungen der beruflichen Vorsorge auf den Arbeitsmarkt (Schaetti 1990).

Im Bereich der *multivariaten Analyse* wird hier nur gerade die multiple Regressionsanalyse erläutert. Für die übrigen Verfahren sei auf die Literatur verwiesen (Backhaus et al. 1994; Log-lineare Modelle: Fienberg 1980; Knoke und Burke 1980; Fingleton 1984; Ishii-Kuntz 1994; für Anwendungen vgl. Keller 1991; Widmer und Buri 1992; Latente Klassenanalyse (LCA): McCutcheon 1987, für eine Anwendung siehe Keller 1991; Lineare-Strukturgleichungsmodelle (LISREL): Bollen 1989; Bollen und Long 1993; Kühnel 1993; Saris und Stronkhorst 1984). Die multivariate (oder multiple) Regressionsanalyse dehnt das bivariate Regressionsmodell weiter aus, indem nicht nur eine unabhängige Variable (x), sondern mehrere exogene Faktoren (x_1, x_2, \dots, x_n) in das Modell integriert werden. In der Evaluationsforschung können damit Analysen durchgeführt werden, welche nicht nur den Zusammenhang zwischen einer staatlichen Massnahme (bspw. x_1) und einer Zieldimension (y) untersuchen, sondern auch Dritt-Variablen (x_2, x_3, \dots, x_n), die ebenfalls einen Einfluss auf die Zieldimension ausüben könnten, in die Analyse zu integrieren (vgl. Kapitel 3.3.). Das einfachste Regressionsmodell geht von einer relativ grossen Anzahl von Annahmen aus, welche die Anwendungsmöglichkeiten in sozialwissenschaftlichen Evaluationsforschungen deutlich einschränken. Hinzuweisen ist hier auch auf das Erfordernis, dass zur Schätzung von Regressionsgleichungen eine ausreichende Zahl von Beobachtungen zur Verfügung steht. Die Zahl der Freiheitsgrade (df) muss im absoluten Minimum 20 betragen, was bedeutet, dass für ein Modell mit gerade drei exogenen Faktoren (x_1, x_2 und x_3) mindestens 24 Fälle zur Verfügung stehen müssen. Ein Beispiel für eine Anwendung des multivariaten Regressionsmodells in der Evaluation findet sich in der Evaluationsstudie zu den technologischen Auswirkungen der umweltpolitischen Vorschriften im Bereich der Hausfeuerungen, die durch Andreas Balthasar und Carlo

Knöpfel (1994: 99-114) erarbeitet wurde. Weitere evaluative Anwendungen finden sich in zwei Arbeiten zur Wirkungsüberprüfung der Kehrtrittsgebühren (Hausheer 1991; Schwarz und Brozovic 1992).

Bei *Längsschnittanalysen* sind im Prinzip dieselben Verfahren wie bei der Querschnittanalyse einsetzbar, wobei sich verschiedene der diskutierten Verfahren besser im Querschnittbereich umsetzen lassen. Die Verfahrensauswahl wird im Bereich von Längsschnittanalysen durch einige, eigens für Untersuchungen dieser Art entwickelte Verfahren ergänzt. Die Analyse in der Längsschnittdimension beinhaltet zusätzlich zu den Problemen, die auch im Querschnittbereich auftreten, weitere Schwierigkeiten (siehe dazu ausführlicher Menard 1991). An erster Stelle ist dabei darauf hinzuweisen, dass es sich bei Daten, die entlang der Zeitachse orientiert sind, um Beobachtungen handelt, welche nicht unabhängig voneinander sind. So besteht oftmals ein sehr starker Zusammenhang von Datenwerten zum Zeitpunkt t mit den Angaben zum vorhergehenden Zeitpunkt ($t-1$) oder zur nachfolgenden Beobachtung ($t+1$). Diesem Umstand können viele der oben beschriebenen Querschnittverfahren zu wenig Rechnung tragen. Diese gehen oft davon aus, dass die Daten aus voneinander unabhängigen Beobachtungen stammen. Bei einem Verstoss gegen die Annahmen führen diese Verfahren zu verzerrten Resultaten. Abbildung 9 enthält eine Zusammenstellung einiger Analyseverfahren zur Längsschnittanalyse. Auch hier ist deutlich darauf hinzuweisen, dass die in der Tabelle enthaltene Zusammenstellung nur eine kleine Auswahl der möglichen Verfahren darstellt.

Abbildung 9 Datenerfordernisse ausgewählter Verfahren in Längsschnittanalysen

Verfahren	Anzahl Datenpunkte	Messniveau	Fehlende Werte
<i>graphische Analyse</i>	frei	frei	möglich
<i>Vorher/Nachher-Vergleich</i>	mind. 1 Prä- und 1 Postintervention	frei	möglich
<i>Lebensverlaufs-analyse</i>	vollständiger Lebensverlauf	kategorial (nominal & ordinal)	nicht möglich
<i>Regressionssätze</i>	über 20 Datenpunkte (df min. 20)	kardinal/ (kategorial)	problematisch
<i>Box-Jenkins-Ansätze</i>	mind. 30 Prä- und 30 Postintervention	kardinal	problematisch
<i>Pooled Time Series</i>	$N \times T$ über 30	kardinal/ (kategorial)	problematisch

Die Tabelle enthält zu jedem der genannten Verfahren die Anforderungen, die an das Datenmaterial gestellt werden: die Anzahl der Datenpunkte, welche für die Analyse zur Verfügung stehen müssen, das erforderliche Messniveau der abhängigen Variable (y), also der Zieldimension, sowie die Flexibilität des Verfahrens in Hinblick auf den Umgang mit fehlenden Werten im Datenmaterial.

Ambivalent einsetzbar und für jede weitergehende Analyse unverzichtbar ist die optische Auswertung der vorhandenen Daten. Aufgrund einer *graphischen Auswertung* sind vielfach schon interessante Erkenntnisse möglich. Bei der graphischen Darstellung von Daten ist darauf zu achten, dass die zur Anwendung gelangenden Massstäbe keine verzerrende Wirkung aufweisen. Insbesondere ist wenn immer möglich auf logarithmische Skalen und auf die Aussparung von Skalenteilen zu verzichten. Anwendungsbeispiele zu diesem Verfahren finden sich in Evaluationsstudien sehr häufig. Exemplarisch sei hier auf die Studie zur Wohneigentumsförderung

des Bundes hingewiesen, welche diese Technik einsetzt (siehe Schulz, Muggli und Hübschle 1993).

Vorher/Nachher-Vergleiche mit oder ohne Kontrollgruppe(n) sind in Evaluationen häufig anzutreffen (siehe etwa INFRAS 1990). Aufgrund der geringen Datenerfordernisse - es muss lediglich eine Angabe vor der Einführung einer Massnahme und eine für die Zeit nach der Einführung vorliegen - ist auch dieses Verfahren oft verwendbar. Es besteht die Möglichkeit, eine Vorher/Nachher-Untersuchung mittels statistischer Analysen durchzuführen. Dabei stehen Verfahren im Vordergrund, welche testen, inwiefern die Differenzen von der Vor- zur Nachinterventionsperiode signifikant sind. Diese Analysen können sowohl das Niveau (Mittelwertvergleiche) als auch die Streuung (Varianzvergleiche) der abhängigen Variablen über die Zeit hinweg betreffen. Problematisch sind derartige Untersuchungen vor allem dann, wenn nur wenige Datenpunkte vorliegen. Insbesondere bei stark volatilen Datenreihen oder wenn ein Trend vorliegt, kann diese Untersuchungsanlage zu verzerrten Resultaten führen. Empfehlenswert zur Erhöhung der Reliabilität und Validität ist sicherlich immer der Einbezug von Kontrollgruppen in unterschiedlichster Form (vgl. dazu die ausführliche Darstellung in Cook und Campbell 1979: 95-146; zur Analyse vgl. Reichardt 1979).

Die *Lebensverlaufsanalyse* (Ereignis(daten)analyse, "event history analysis") ermöglicht die Analyse von Längsschnittdaten mit kategorialer abhängiger Variable. Verbreitet ist diese Untersuchungsform bisher vor allem im Bereich der quantitativ ausgerichteten, soziologischen Biographieforschung. Für weitere Hinweise sei auf die Literatur verwiesen (Blossfeld, Hamerle und Mayer 1986; Diekmann und Mitter 1984; Yamaguchi 1991).

Der *Regressionsansatz* ist nicht nur bei Querschnitt-, sondern auch in Längsschnittuntersuchungen einsetzbar. Für die Schätzung von Zeitreihenregressionen kann jedoch zumeist nicht auf die OLS-Schätzprozeduren zurückgegriffen werden, sondern es sind andere Verfahren zu verwenden. In der Ökonometrie ist eine Vielzahl verschiedener Techniken zur unverzerrten Analyse von Zeitreihendaten entwickelt worden. Diese setzen jedoch zumeist ein tiefergehendes Verständnis der Schätzmodelle voraus, die in der Literatur ausführlich erläutert sind (Ostrom 1990; Gujarati 1988; Stewart 1991; Johnston 1972). Als Anwendungsbeispiel hierzu sei auf die Evaluationsstudie zu den Auswirkungen der Ausländerpolitik auf den Strukturwandel in der Schweiz hingewiesen (Stalder, Frick, Gaillard und Salzgeber 1994).

Ein reines Zeitreihenverfahren, das kein Äquivalent im Querschnittbereich kennt, stellt die Interventionsanalyse im Rahmen der *Box-Jenkins-Ansätze* dar. Diese, auch unter der Bezeichnung ARIMA-Modelle (autoregressive, integrierte, moving-average-Modelle) bekannten Analyseverfahren gehen von den in Zeitreihen enthaltenen Strukturen, wie Trends, Zyklen und Saisonalitäten aus, und versuchen so die Effekte einer Intervention zu erfassen (vgl. dazu Box und Jenkins 1976; McCleary und Hay 1980; Schmitz 1989; Wei 1990). Diese Methodik kann sowohl in bivariater (Interventionsanalyse) wie auch in multivariater Form (Transferfunktionsanalyse) eingesetzt werden. Der grosse Vorteil dieser Methode besteht darin, dass auch komplexere Interventionen und Wirkungsverläufe erfasst werden können und dass, im Gegensatz zu anderen Verfahren, die spezielle Charakteristik von Zeitreihendaten beachtet wird. Nachteilig wirken sich vor allem zwei Faktoren aus, die den Einsatz dieses Verfahrens oft verhindern. Erstens sind die Erfordernisse, welche an das empirische Datenmaterial gestellt werden, sehr hoch (vgl. Abbildung 9). Zweitens ist das Verfahren in der Umsetzung doch recht anspruchsvoll und zeitintensiv. Als Beispiele für die Anwendung dieses Verfahrens sei auf eine Studie zur Evaluation luftreinhaltepolitischer Massnahmen in der Schweiz (Widmer 1991) verwiesen.

Abbildung 9 enthält ein weiteres Verfahren, das jedoch kein reines Längsschnittmodell darstellt. Es handelt sich dabei um die *"pooled time series"*, ein Verfahren, das eine kombinierte Längsschnitt-Querschnittanalyse ermöglicht (vgl. dazu Stimson 1985; Sayrs 1989). Aufbauend auf dem Regressionsansatz hat dieses Verfahren den Vorteil, dass in Situationen, in denen nur in einer Vergleichsdimension keine befriedigende statistische Analysemöglichkeit bestehen würde, trotzdem eine statistische Analyse ermöglicht wird. Sehr häufig zeigt sich nämlich in der Praxis das Problem, dass keine ausreichende Zahl von Fällen (im Querschnitt) oder von Beobachtungszeitpunkten (im Längsschnitt) zur Verfügung stehen, um z.B. eine Regressionsanalyse durchzuführen. Durch eine Kombination des Datenmaterials der beiden Vergleichsdimensionen ist die Zahl der Beobachtungen jedoch ausreichend gross, um eine zuverlässige Schätzung zu ermöglichen. Auch diese Methodik ist in der Umsetzung nicht immer einfach. Durch die dimensionale Erweiterung der Analyse erhöht sich auch deren Komplexitätsgrad. Ein Beispiel für die Umsetzung findet sich in einem Artikel, der sich mit den Auswirkungen der Arbeitslosigkeit auf den Wertewandel befasst (Clarke und Dutt 1991).

Die Vergleiche in der Zeitdimension haben, trotz allen Schwierigkeiten bei deren Implementation, einige gewichtige Vorteile (vgl. dazu Menard 1991: 5-21), die gerade in der Evaluationsforschung von grösstem Interesse sein können. Erwähnt sei an dieser Stelle etwa die analytische Trennung von Alters- und Zeiteffekten, welche nur in der Zeitdimension zu realisieren ist, die Bedeutung der zeitlichen Ordnung bei der Analyse von Wirkungsketten oder die Dynamik sozialer Prozesse, die mit Querschnittanalysen nicht ausreichend erfasst werden kann. Die Evaluationsforschung untersucht zumeist Veränderungshypothesen, und Veränderungen sind wohl am zweckmässigsten im Zeitverlauf analysieren.

Die hier vorgestellten Analyseverfahren stellen sehr unterschiedliche Anforderungen an das vorhandene Datenmaterial und an die Kenntnisse des Analytikers. Weiterhin ist die Verfahrensauswahl auf die Fragestellung und auf den Themenbereich abzustimmen. Es ist im Auge zu behalten, dass Analysetechniken keineswegs neutrale Werkzeuge darstellen. Praktisch alle quantitativen Verfahren gehen von mehr oder weniger engen Prämissen aus, welche in der Interpretation nicht vergessen gehen sollten. Evaluationen entstehen häufig unter engen zeitlichen und finanziellen Rahmenbedingungen. Aus dieser Konstellation heraus besteht die Gefahr, dass Analyseverfahren nicht sachgerecht eingesetzt werden oder dass die Ergebnisse der Analysen über- oder gänzlich falsch interpretiert werden. Beherrschen die Evaluatoren oder die Evaluatorinnen eine Verfahrenstechnik, sind sie versucht, diese Kompetenz auch in nachfolgenden Evaluationen einzusetzen, auch wenn dies von der Sache her wenig zweckmässig erscheint. Dadurch kann die Verfahrenstechnik zum Selbstzweck werden. Die Wahl der Evaluierenden bestimmt damit die eingesetzte Methodik und nicht - wie dies zu wünschen wäre - die zu untersuchende Fragestellung. Andererseits ist auch ein Evaluationsteam nicht in der Lage, alle dargestellten Verfahren (und die grosse Zahl darüber hinaus bestehender Techniken) zu beherrschen. Weil kaum Zeit besteht, sich für eine spezifische Evaluation in eine gänzlich neue Methodik einzuarbeiten und diese dann auch noch in hoher Qualität durchzuführen, scheint es oftmals angebracht, sich das nötige Know-how durch Beizug einer Expertenperson zu beschaffen.

Zum Abschluss seien noch vier Empfehlungen formuliert:

- Die Qualität des Methodeneinsatzes ist entscheidend. Im Zweifelsfall scheint es deshalb empfehlenswert, sich für die "einfachere" Analyse-

technik zu entscheiden und sich dabei um eine möglichst hochstehende Umsetzung zu bemühen.

- Sehr oft bauen quantitative Datenanalyseverfahren aufeinander auf, sind also hierarchisch organisiert. Es empfiehlt sich, nicht direkt die komplexeste Analysetechnik anzusteuern, sondern zuerst die zugrundeliegenden Verfahren zu verwenden, um anschliessend die vorgesehene komplexe Methodik umzusetzen. Beispielsweise ist es von Vorteil, zuerst eine gründliche univariate Analyse des Datenmaterials vorzunehmen, bevor man sich auf ein multivariates Verfahren konzentriert.
- Lassen die Rahmenbedingungen dies zu, ist es zweckmässig, die Evaluation nicht nur auf eine Auswertungsmethodik auszurichten, sondern verschiedene Analyseverfahren einzusetzen.
- Bei der Formulierung der Schlussfolgerungen einer Evaluation sollte man sich immer vergewissern, ob man dabei alle vorgenommenen Annahmen berücksichtigt und alle empirischen Unzulänglichkeiten, welche jede Evaluationsstudie enthält, einbezogen hat.

12.3.3. Darstellung der Ergebnisse

Die Evaluationsforschung hat verschiedene Qualitätskriterien entwickelt, welche Normen zur Durchführung und zur Darstellung evaluativer Untersuchungen enthalten (siehe z.B. Joint Committee 1994; vgl. dazu auch Widmer 1996). Im Zusammenhang mit der Darstellung der Anwendung quantitativer Verfahrenstechniken und deren Ergebnisse sind dabei vor allem die folgenden Aspekte zu beachten:

- *Klarheit des Berichts:* Evaluationsteams sollten darauf bedacht sein, den Evaluationsbericht in einer Form abzufassen, die es erlaubt, die verwendeten Verfahren zu identifizieren und einzuschätzen. Die verwendeten Techniken müssen detailliert und eindeutig beschrieben werden. Zur Entlastung des Berichts kann dies auch in einem methodischen Anhang geschehen. Es sollte das Ziel eines Evaluationsberichts sein, die durchgeführten Analysen für den Leser nachvollziehbar auszugestalten. Dies bedingt, dass Entscheide, welche im Untersu-

chungsverlauf getroffen werden, nicht nur dargelegt, sondern auch explizit begründet werden.

- *Offenheit in der Darlegung:* Bei der Darstellung der Evaluation sollten auch Probleme oder Schwierigkeiten, die im Evaluationsprozess entstanden, zur Sprache kommen. Insbesondere gilt es auch auf die Konsequenzen derartiger Aspekte hinzuweisen. Ebenso ist es unumgänglich, dass der Evaluationsbericht klare Hinweise auf die Grenzen der Aussagekraft der Evaluation enthält.
- *Begründung der Folgerungen:* Der Evaluationsbericht sollte es dem Leser ermöglichen, einen Bezug zwischen empirischer Analyse und angebrachten Schlussfolgerungen herzustellen. Enthält eine Evaluation weitergehende Folgerungen und Empfehlungen, welche sich nicht in direkter Weise mit den empirischen Analysen begründen lassen, ist dies deutlich hervorzuheben.

Diese Hinweise gelten natürlich nicht nur für schriftliche Evaluationsberichte, sondern sind in analoger Weise auch bei einer Berichterstattung in anderer Form (mündliche Berichterstattung, Video etc.) zu beachten (vgl. dazu Morris, Fitz-Gibbon und Freeman 1987; Wottawa und Thierau 1990: 156-164; Passow 1987).

Die Anwendung quantitativer Verfahren im Bereich der Evaluationsforschung steht in einem Spannungsfeld zwischen methodischer Präzision und Anschlussfähigkeit bei den Adressaten. Gemeint ist damit die Problematik, dass höher entwickelte und vielleicht adäquatere oder aussagekräftigere statistische Verfahren gegenüber der Praxis schwerer mitteilbar werden (vgl. Kapitel 10). Daraus ergibt sich die Frage, inwiefern Evaluationsverfahren durch die Adressaten überhaupt nachvollziehbar sein sollen. Hier bestehen in der Literatur unterschiedliche Auffassungen. Für das Primat der Anschlussfähigkeit sprechen teilweise die Erkenntnisse aus der Forschung zur Nutzung von Evaluationsergebnissen oder von sozialwissenschaftlichem Wissen im allgemeinen (siehe etwa Alkin 1990: 19-78; Beck und Bonss 1989; Wittrock 1991; Chen 1994; Shadish 1994; und natürlich: Weiss 1981; 1987; 1993). Dagegen sprechen die Erfahrungen aus anderen, nicht-sozialwissenschaftlichen Disziplinen mit anwendungsorientierter Charakteristik. Aber auch dort wird vermehrt eine "Kontrolle" der Forschungstätigkeiten eingefordert.

Evaluationsberichte haben sich in jedem Fall darum zu bemühen, den Prozess, die Prozeduren und Verfahren und die Ergebnisse einer Evaluation möglichst klar, offen und nachvollziehbar darzulegen, also Transparenz zu schaffen. Nur so kann auch erreicht werden, dass der Evaluation die ihr zukommende Beachtung geschenkt wird. Dadurch wird weiter auch die Möglichkeit, dass Evaluationen missbraucht werden, eingeschränkt. Die Qualität der Berichterstattung ist sowohl durch sprachliche, formale, als auch durch inhaltliche und methodische Aspekte geprägt. Die Möglichkeiten für eine übersichtliche Gestaltung des Berichts (Stichworte: graphische Darstellungen, tabellarische Übersichten, Verzeichnisse, Querverweise usw.) sollten genutzt werden. Die Anfertigung eines qualitativ hochstehenden Evaluationsberichts braucht Zeit und Sachkenntnis. Dafür sind entsprechende Mittel einzuplanen. Auch die gekonnteste Evaluation ist wenig hilfreich, wenn es nicht gelingt, die Evaluation (und nicht nur ihre Ergebnisse) zu kommunizieren.